Study of Extreme Magnetopause Distortions under Varying Solar Wind Conditions

Niklas Grimmich[1], Ferdinand Plaschke[1], Martin O Archer[2], Daniel Heyner[1,3], Johannes Z D Mieth[1], Rumi Nakamura[3,4], and David G Sibeck[4]

[1]Institut für Geophysik und Extraterrestrische Physik, Technische Universität Braunschweig
[2]Department of Physics, Imperial College London
[3]Space Research Institute, Austrian Academy of Sciences
[4]NASA Goddard Space Flight Center

July 20, 2023

Study of Extreme Magnetopause Distortions under Varying Solar Wind Conditions

Niklas Grimmich[1], Ferdinand Plaschke[1], Martin O. Archer[2], Daniel Heyner[1], Johannes Z. D. Mieth[1], Rumi Nakamura[3], David G. Sibeck[4]

[1]Institut für Geophysik und Extraterrestrische Physik, Technische Universität Braunschweig, Braunschweig, Germany
[2]Department of Physics, Imperial College London, London, UK
[3]Space Research Institute, Austrian Academy of Sciences, Graz, Austria
[4]NASA Goddard Space Flight Center, Greenbelt, Maryland, US

Key Points:

- More than 160.000 magnetopause crossings (MPCs) identified in THEMIS data between 2007 and 2022 using a Random Forest Classifier
- MPCs that extremely deviate in location from the Shue et al. (1998) model are quite common
- Important solar wind parameters associated with deviations include the IMF cone angle, solar wind velocity and Alfvén Mach number

Corresponding author: Niklas Grimmich, `n.grimmich@tu-braunschweig.de`

Abstract

To first order, the magnetopause (MP) is defined by a pressure balance between the solar wind and the magnetosphere. The boundary moves under the influence of varying solar wind conditions and transient foreshock phenomena, reaching unusually large and small distances from the Earth. We investigate under which solar wind conditions such extreme MP distortions occur. Therefore, we construct a database of magnetopause crossings (MPCs) observed by the THEMIS spacecraft in the years 2007 to mid-2022 using a simple Random Forest Classifier. Roughly 7% of the found crossing events deviate beyond reported errors in the stand-off distance from the Shue et al. (1998) MP model and thus are termed extreme distortions. We find the occurrence of these extreme events in terms of expansion or compression of the MP to be linked to different solar wind parameters, most notably to the IMF magnitude, cone angle, velocity, Alfvén Mach number and temperature. Foreshock transients like hot-flow anomalies and foreshock bubbles could be responsible for extreme magnetospheric expansions. The results should be incorporated into future magnetopause models and may be helpful for the reconstruction of the MP locations out of soft x-ray images, relevant for the upcoming SMILE mission.

1 Introduction

Earth's magnetopause is the boundary layer between the solar wind and the terrestrial magnetosphere. It is an obstacle for the incoming super-magnetosonic solar wind. A bow shock (BS) upstream of the MP decelerates the solar wind and then deflects the plasma around the magnetosphere. The region between the magnetopause and the bow shock is called magnetosheath (e.g., Baumjohann & Treumann, 1997). Depending on the angle between the interplanetary magnetic field (IMF) vector and the bow shock normal, the respective bow shock region (and the magnetosheath) may be denoted as quasi-parallel (angle $< 45°$) or quasi-perpendicular (angle $> 45°$). Upstream of the quasi-parallel bow shock, an extended foreshock region can form, permeated by waves which are excited due to the interaction of the solar wind with particles reflected at and back streaming from the BS (e.g., Eastwood et al., 2005).

Dynamical changes in the solar wind and subsequently in its interaction with the BS influence the magnetosheath flow and impact the MP location and shape. In the absence of reconnection, when the MP can be described as a rotational discontinuity, the MP is well-characterized as a tangential discontinuity at which pressure balance should hold. On the magnetospheric side, the magnetic pressure is the most important contributor to that balance, while on the magnetosheath side dynamic, plasma (thermal) and magnetic pressures (from the draped IMF) contribute significantly (e.g., Shue & Chao, 2013). Thus, variations of the total pressure in the solar wind and in the magnetosheath lead to inward and outward motion of the MP. Additionally, strong southward IMF conditions lead to magnetic flux erosion from the dayside MP via magnetic reconnection and therefore inward motion of the dayside MP (Aubry et al., 1970; Sibeck et al., 1991; Shue et al., 1997, 1998). Solar wind dynamic pressure, IMF strength and orientation and the dipole tilt angle can be identified as the parameters influencing the MP location (Sibeck et al., 1991; Shue et al., 1997; Liu et al., 2012). Consequently, many empirical MP models use the solar wind dynamic pressure p_{dyn}, the IMF B_z-component and in some instances the dipole tilt as input parameters (e.g., Fairfield, 1971; Sibeck et al., 1991; Shue et al., 1997; Chao et al., 2002; Lin et al., 2010; Nguyen et al., 2022c, and many others). In these models, the MP stand-off distance R_0 serves as an indicator for the overall location of the boundary layer. R_0 is often strongly dependent on one or both of the two parameters p_{dyn} and B_z.

Newer models like the one from Lin et al. (2010) or Nguyen et al. (2022c) use additional parameters like the solar wind magnetic pressure and the dipole tilt to take asymmetries and cusp indentation into account, enhancing the forecasting accuracy of the model, e.g., shown by Case and Wild (2013) for the Lin et al. (2010) model. Physics-based MHD models like, e.g., Liu et al. (2015) include all IMF components, P_{dyn} and the dipole tilt as parameters to give an even better forecasting accuracy under normal solar wind conditions. Nevertheless, most models fail to predict magnetopause locations under extreme pressure conditions (e.g., Tátrallyay et al., 2012; Suvorova & Dmitriev, 2015). In these cases, other parameters can become more significant. One of those parameters, which to our knowledge is not included in the models and also describes the IMF orientation, is the IMF cone angle $\vartheta_{\mathrm{cone}}$ between the Earth-Sun-line and the IMF vector. Magnetospheric expansions beyond the magnetopause model predictions are often found when the IMF is quasi-radial ($\vartheta_{\mathrm{cone}} < 30°$) (Fairfield et al., 1990; Suvorova et al., 2010; Dušík et al., 2010; Samsonov et al., 2012; Grygorov et al., 2017). Another parameter could be the IMF clock angle $\vartheta_{\mathrm{cone}}$ between IMF B_y- and B_z-components, as Lu et al. (2013) showed in global MHD simulations, that the IMF B_y and B_z components might have influence on the MP shape.

In addition to changes in the dynamic pressure and/or IMF orientation, other phenomena have been discussed as origins of MP disturbances, which can lead to extreme R_0 values. Phenomena originating near the magnetopause include magnetic reconnection and associated flux transfer events (FTE, e.g., Elphic, 1995) or the Kelvin-Helmholtz instability (KHI, e.g., Johnson et al., 2014). In the magnetosheath, so called magnetosheath or high-speed jets (HSJs) can travel from their point of origin at the bow shock down to the magnetopause and cause an indentation and excitation of surface waves (Shue et al., 2009; Plaschke et al., 2018; Archer et al., 2019). Finally, kinetic transients in the foreshock region, like hot-flow anomalies (HFAs) or foreshock bubbles (FBs) and ULF-wave generated phenomena like foreshock cavitons, short large-amplitude magnetic structures (SLAMS) or shocklets, can impact the MP in different ways as well (Sibeck et al., 1999; Jacobsen et al., 2009; Turner et al., 2011; Archer et al., 2015; H. Zhang et al., 2022). Some of these phenomena only result in localized distortions (e.g. HFAs, Sibeck et al., 1999; Turner et al., 2011), others could have global impacts (e.g. FBs, Archer et al., 2015).

These phenomena and the solar wind-magnetosphere interactions have been studied for two decades using data from several multi-spacecraft missions. Cluster (Escoubet et al., 2001) contributed significantly to the exploration of different plasma regions of the magnetosphere, advancing our understanding of reconnection and the movement of the magnetopause (see Haaland et al., 2021, for a comprehensive overview). The Time History of Events and Macro-scale Interactions during Substorms (THEMIS) mission (Angelopoulos, 2008) enabled observations of solar wind phenomena and direct responses in the magnetosphere due to the special orbit configuration of the multiple spacecraft. The aim of the most recent mission MMS is to study in detail magnetic reconnection at the smallest scales (Burch et al., 2016).

Typically, all these spacecraft can only observe the MP at the position and time they cross this boundary or when the MP is in motion and moves over the spacecraft. So far, global observations of the MP have not been possible.The upcoming Solar Wind Magnetosphere Ionosphere Link Explorer (SMILE) mission will provide the first opportunity to observe the location, shape and motion of the dayside MP at any given time (Raab et al., 2016; Branduardi-Raymont et al., 2018), based on measurements of soft x-rays. Soft x-rays are emitted during solar wind charge exchange with neutrals from the Earth's exosphere (e.g., see review by Kuntz, 2019). Studies of this phenomenon in the near-Earth regions showed the possibility to image the magnetospheric boundary layers in soft x-ray wavelengths and reconstruct the magnetopause surface from the images (e.g., Sibeck et al., 2018; C. Wang & Sun, 2022). SMILE will take advantage of this to study the whole dayside magnetosphere from a polar orbit and image the soft x-rays with a Soft X-ray Imager (SXI) to track the magnetopause motion on global scales. Additional instrumentation of SMILE will include a Magnetometer (MAG), a Light Ion Analyser (LIA) and an Ultra-Violet Imager (UVI) which will monitor the plasma environment, in particular the solar wind conditions, and the auroral oval in UV wavelengths, respectively. Thereby, the motion of the magnetopause can be linked to the upstream plasma conditions and the ionospheric response.

The SMILE mission is expected to launch in late 2024 or early 2025. In preparation, much effort is put into the development of MP reconstruction techniques based on simulated SXI images, for which fundamental knowledge about the magnetopause shape and behaviour is needed (see C. Wang & Sun, 2022, and references therein). The influence of IMF parameters on it has been subject to several statistical studies (e.g., Plaschke, Glassmeier, Auster, Angelopoulos, et al., 2009; Dušík et al., 2010; Staples et al., 2020; Nguyen et al., 2022b). In this study, we focus on extreme MP locations, which cannot be explained with a common MP model like the improved Shue et al. (1998) model, theoretically capable of predicting the MP location under extreme solar wind conditions. The reason for this is most likely due to the fact that such models are designed to be optimal around the typical conditions and give an average location and shape of the MP. Therefore, extreme and unusual conditions are given less weight in fitting the models, resulting in model predictions deviating under such conditions.

In previous works extreme MP locations were often only studied on a case by case bases. Utilizing multiple years of THEMIS data we can construct a large database of crossing events and study the extreme and unusal MP locations with a statistical approach. In particular, we are interested under which solar wind conditions these events occur. Understanding the origins of extreme MP locations, which might be parameters included in other models or extreme conditions, will help improve future models and help understand the interaction between the solar wind and the magnetosphere in more detail. In the following, we give a brief introduction to the used spacecraft data (section 2). We describe the construction of our database of magnetopause crossings observed by the THEMIS spacecraft (section 3) and show the results (section 4). We then compare the solar wind conditions for which extreme events occur with the standard solar wind distributions (section 5). Eventually, we discuss and summarize our findings (section 6).

2 Spacecraft Data

Since 2007 the spacecraft of the THEMIS mission have been orbiting Earth near the equatorial plane to investigate the plasma environment in the near-Earth region (Angelopoulos, 2008). For the identification of MPCs in the timespan of 2007 to 2022, we use the magnetic field data from the Fluxgate Magnetometer (FGM, Auster et al., 2008), and particle data and moments from the Electrostatic Analyzer (ESA, McFadden et al., 2008). Data from the entire 15 years interval are used from probes THA, THD, and THE, while THB and THC only contribute data until the end of 2009, as they were then sent into lunar orbits, becoming the ARTEMIS mission (Angelopoulos, 2011).

FGM and ESA data are used in the spin-resolution (FGM) and reduced mode (ESA) with cadences of about 3 to 4 s. Low resolution FGM data and full mode ESA data are used to bridge bigger data gaps (> 15 min) (see data treatment by Nguyen et al., 2022a). This occurs almost exclusively in the velocity data of ESA, leading to some uncertainties in this data which can be compensated in our detection method. The FGM and ESA data are synchronized and resampled to common 3 s time stamps. Finally, we average the data in a moving 60 s window for each time step, to smooth out turbulent fluctuations which could be misidentified as MPCs. The data is processed in 1-hour intervals with an overlap of 2 minutes into the next interval. Intervals were omitted for data gaps that could not be bridged, i.e. if less than 15 mins of data were available in both the high and low resolution. This is necessary, as large data gaps lead to jumps in the data which could be misinterpreted as MPCs. Results are combined to a bigger dataset afterwards.

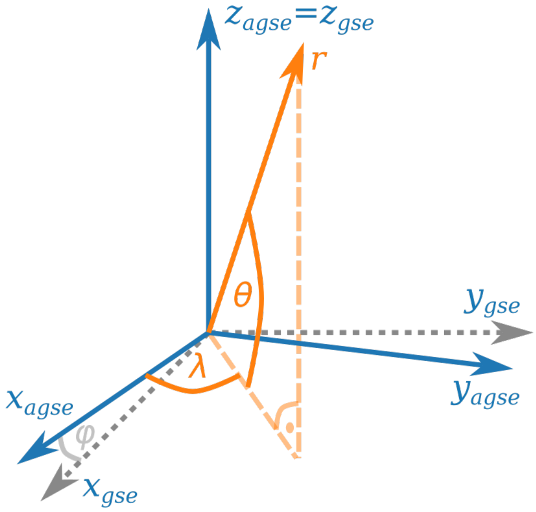

Figure 1. Orientations and relations of the two main coordinate systems. The grey axes depict the standard GSE (geocentric solar ecliptic) system. The cartesian and spherical AGSE (aberrated geocentric solar ecliptic) coordinate axes are depicted in blue and orange, respectively.

All vector quantities in the dataset are transformed into the AGSE (aberrated geocentric solar ecliptic) coordinate system with an average aberration angle of $\varphi \sim 4.3°$ resulting from the Earth's orbital velocity of 30 km/s around the Sun and an average solar wind speed of 400 km/s. Taking this aberration effect on the MP into account, improves the prediction of MP models (e.g., Safránková et al., 2002).

We limit our investigation to the dayside magnetosphere (position in AGSE x-direction larger than 0) outside the nominal plasmasphere (minimal radial distance from the Earth's center larger than 4 R_{E}). These conditions have to be fulfilled throughout the 1-hour intervals. This can result in fewer observations near the terminator.

In addition to the observational data, we construct a dataset containing the orbital data of the THEMIS probes in the cartesian (x, y, z) and the spherical AGSE coordinates $(r, \theta, \lambda$, see Fig. 1) in 1-minute resolution for the all selected time intervals. This dataset also comprises the equivalent stand-off distance $R_{0,\mathrm{sc}}$ calculated with the Shue et al. (1997, 1998), hereafter SH98, model equation, as done in previous studies (Plaschke, Glassmeier, Auster, Constantinescu, et al., 2009; Plaschke, Glassmeier, Auster, Angelopoulos, et al., 2009; Staples et al., 2020):

$$R_{0,\mathrm{sc}} = r\left(\frac{2}{1+\cos\zeta}\right)^{-\alpha}. \tag{1}$$

Here r is the radial distance from the Earth's center to the spacecraft and ζ is the zenith angle between the x-axis and the Earth-spacecraft-line (denoted by θ in Shue et al., 1997, 1998). The flaring parameter α is calculated with the formula given by Shue et al. (1998), using the appropriate dynamic solar wind pressure p_{dyn} and IMF component $B_{z,\mathrm{IMF}}$ for all orbital points:

$$\alpha = \left(0.58 - 0.007\frac{B_{z,\mathrm{IMF}}}{\mathrm{nT}}\right)\left[1 + 0.024\ln\left(\frac{p_{\mathrm{dyn}}}{\mathrm{nPa}}\right)\right]. \tag{2}$$

We take into account that our approximation of a static solar wind speed for the aberration effect results in mean errors of 0.034 R_{E} for $R_{0,\mathrm{sc}}$ and 0.823° for the longitude-position, which have no drastic influence on our study. To calculate the presented errors, we average the differences between position values of THEMIS adjusted with a dynamic solar wind and with a static solar wind aberration.

The appropriate solar wind parameters are obtained from the high resolution 1-min OMNI dataset, which mainly combines the WIND (Lepping et al., 1995; Ogilvie et al., 1995) and ACE (Stone et al., 1998; Smith et al., 1998; McComas et al., 1998) spacecraft data, time-shifted to the bow shock nose (see King & Papitashvili, 2005, for details on the shift technique). Smaller data gaps up to 5 minutes in the OMNI dataset are bridged by linear interpolation.

3 Magnetopause Crossing Identification Method

Our identification process utilizes a combination of supervised machine learning methods and a threshold-based classification, to infer crossing events from automatically labelled data.

Recent studies have already shown the efficiency of classifying the near-Earth regions from spacecraft data with machine learning methods (e.g., Breuillard et al., 2020; Olshevsky et al., 2021; Nguyen et al., 2022a). In particular Nguyen et al. (2022a) showed that even a simple machine learning algorithm like the Gradient Boosting Classifier can outperform manually set threshold based detection methods of the three typical near-Earth regions (solar wind, magnetosheath and magnetosphere), reaching more than satisfying accuracies.

Unfortunately, Nguyen et al. (2022a) only inferred if one MPC is found in a 1-hour interval, finding only a limited amount of MPCs with an uncertain location. This is not suitable for our study, as we can not be certain to infer the right model deviations from their catalogue. We aim to construct a database in which extreme MPCs are clearly identified on smaller timescales and with a clear spacecraft location, which can be used in future studies on extreme MP distortions. Nevertheless, we can use the same approach as Nguyen et al. (2022a) in giving every data point a label according to the near-Earth region it most likely pertains to, and then infer the boundary crossings from the labels. For our study, we only need to distinguish between data points that are in the magnetosphere labelled 1 and data points that are not in the magnetosphere labelled 0, facilitating the identification of magnetopause crossings.

In Fig. 2 we present a flow diagram summarizing our identification process. Detailed description can be found in the following sections.

3.1 Machine Learning Algorithms

For our study, we only need to distinguish between data points that are in the magnetosphere labelled 1 and data points that are not in the magnetosphere labelled 0, facilitating the identification of magnetopause crossings.

Nguyen et al. (2022a) trained their algorithm with data resampled to 1-minute resolution consisting of the magnetic field components $(B_x,\ B_y,\ B_z)$, the ion velocity $(v_x,\ v_y,\ v_z)$, the ion density n_{ion} and the ion temperature T_{ion}. We include the magnitude of magnetic field and velocity as well as a flux index $F_{\mathrm{idx}}(t)$ which describes the omnidirectional energy flux of ions with energies between 10^2 eV and 10^4 eV, where the solar wind and magnetosheath regions are easily identified:

$$F_{\mathrm{idx}}^{1e2,1e4}(t) = \log_{10}\left(\int_{10^2\ \mathrm{eV}}^{10^4\mathrm{eV}} \frac{E_{\mathrm{Flux}}(t)}{\frac{\mathrm{eV}}{\mathrm{cm}^2\cdot\mathrm{s}\cdot\mathrm{sr}}}\mathrm{d}E\right). \tag{3}$$

The index reaches high values of 10 and above if the magnetosheath ion population is observed, otherwise $F_{\mathrm{idx}} < 10$ holds (compare panels (5) and (6) of Fig. 3). This $F_{\mathrm{idx}}(t)$

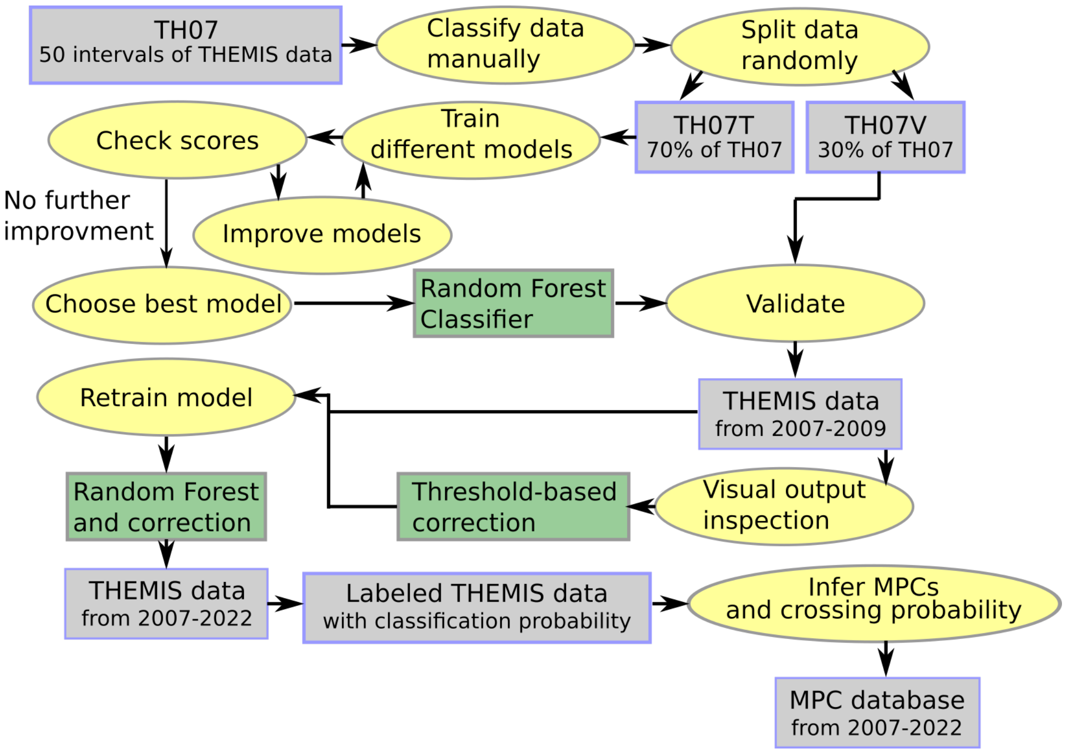

Figure 2. Flow diagram outlining our identification process.

can be better handled by the model than the total energy flux distribution for each time step.

To train and compare different machine learning algorithms, we built a dataset of 50 randomly selected time intervals with different lengths of the initial phase of the THEMIS mission in 2007 (TH07), that represent well outer-magnetospheric dayside observations (see Fig. 4). Each interval contains at least one magnetopause crossing. All data points are then labelled manually by visually inspecting n_{ion} and B_z changes, as well as ion energy flux density measurements, yielding roughly 30,000 labelled data points from inside (Label 1) and outside (Label 0) the magnetosphere with $\sim$ 1300 MPCs for training. Data points in a smeared out MPC or boundary layer are attempted to be separated in the middle of the crossing. Fig. 3 displays one of the intervals from TH07 with all input parameters for the algorithms; it also shows labels given manually and by the trained Random Forest machine learning classifier.

We randomly divide our dataset TH07 into a training set (70% of the data points, TH07T) and a validation set (30% of data points, TH07V). With TH07T we train, test and compare different models to decide which model to utilize for the identification. TH07V is later used to verify the training scores of the best model, assuring the model has not overfitted the trainings data. The nature of our problem, inside (class/label 1) or outside (class/label 0) the magnetosphere, is a binary classification problem which can be tackled with a number of different algorithms (e.g., described in Géron, 2019).

One of the simplest binary classifiers is the Logistic Regression (LR, e.g., Cox & Snell, 1970), predicting the probability of a data point belonging to the positive class (label 1) by calculating a logistic (sigmoid) function of a linear fit of the input data. This algorithm assumes that the data points are linearly distributed in parameter space. Additionally, the data has to be normalized for the algorithm to work properly.

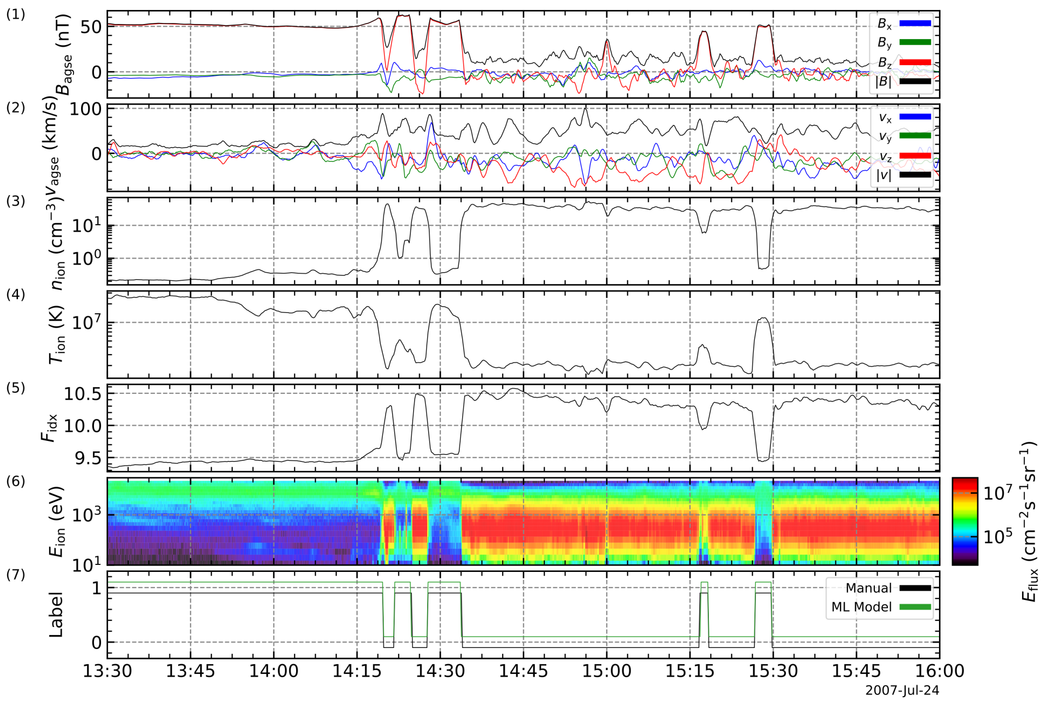

Figure 3. Time series plot of THEMIS data (THE) on the 24 July 2007. From top to bottom the panels display the averaged magnetic field data, the ion velocity, the ion density, the ion temperature, the energy flux density, the flux index and the data label given manually and by a Random Forest Classifier. The label values were shifted slightly for better visual comparison.

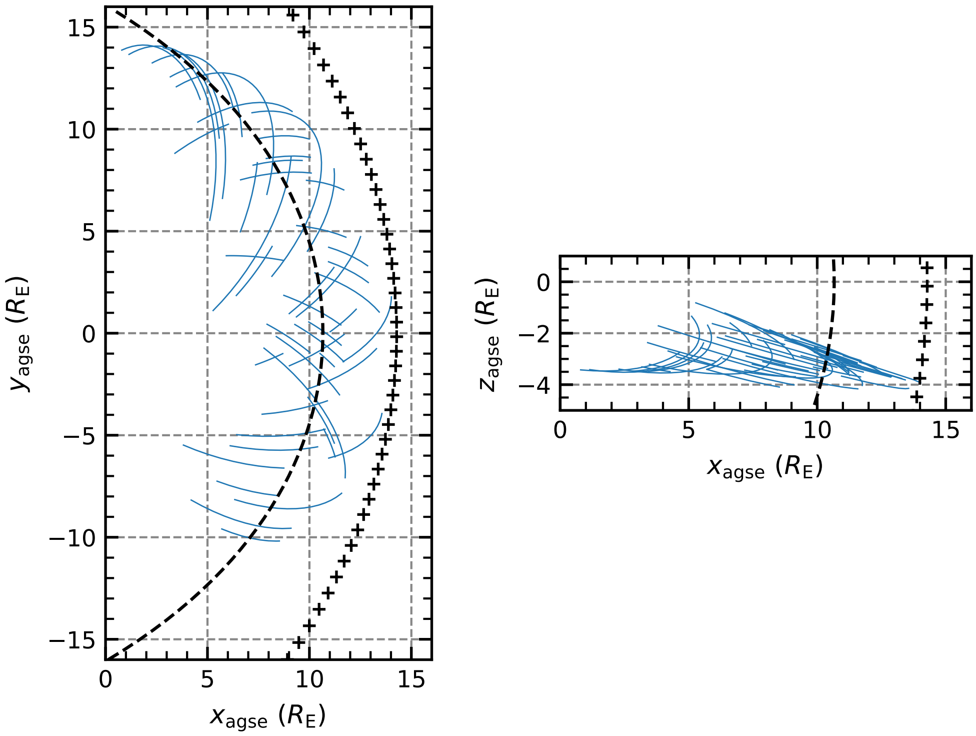

Figure 4. Spatial distribution of 50 training intervals in the AGSE x-y-plane (left panel) and x-z-plane (right panel), respectively. The dashed line represents the Shue et al. (1998) model magnetopause and the black crosses represent the Chao et al. (2002) model bow shock for $B_{z,\mathrm{IMF}} = -1$ nT and $p_{\mathrm{dyn}} = 1.5$ nPa.

Another often used method is the Decision Tree (DT, e.g., Breiman et al., 1984). This algorithm can directly (with only little preprocessing) predict a class from different input data using simple if-then-else decision rules inferred from data features/input parameters. A common problem with DTs, if not restricted correctly, is overfitting, i.e., adapting too tightly to the training data, reducing the adaptability of the model to new data.

More advanced algorithms like the Random Forest (RF, e.g., Breiman, 2001) or Gradient Boosting (GB, e.g., Friedman, 2001) use ensemble methods for their prediction: multiple simple models are trained on the data and the final prediction are then derived from the predictions of all contributing simple models. Both RF and GB algorithms use DTs as basis. The RF algorithm trains a group of DTs on random training data subsets and use the most common prediction in the group as final prediction, therefore reducing the problem of overfitting of the individual DTs. The GB on the other hand sequentially fit DTs on the residual errors of the previously trained DT until the ensemble convergences on the smallest errors, and predicts the class via the sum of the ensembles predictions. These ensemble methods are widely used in many machine learning applications, reaching high accuracies (Géron, 2019). Nguyen et al. (2022a) used the GB algorithm in their work for the identification of the near-Earth regions in spacecraft data.

All the presented algorithms, except the RF, were also compared by Nguyen et al. (2022a). We start the training with more input parameters, hence, we repeat the model comparison here to ensure using the optimal model. For the comparison we have to split our training data TH07T again into training subset (TH07TC) and into a validation subset (TH07TV) with a data ratio of 70/30.

For the first testing round, we utilize the default implementation of the algorithms from Python's Scikit-learn library (Pedregosa et al., 2011) and evaluate the models via the cross validation (CV) scores. Cross validation means that the training data (TH07TC) is split into n equally sized subsets. The model is then trained and evaluated n times with all possible combinations of these subsets as training (n-1 subsets) and validation data (1 subset). Thus, the CV scores give us a mean accuracy (fraction of correct predictions) and standard deviation over all n subsets, working as an indicator for the independence of the data split into training and validation data. Here we utilize a 10-folded CV, i.e., we split the TH07TC into n =10 subsets. Based on this first CV, we can already conclude that the two ensemble classifiers perform better. Nevertheless, as suggested by Géron (2019), we aim at improving all the models by adjusting some important hyperparameters (specific boundaries for the algorithms) using a grid search method: We train and evaluate the models via CV with different parameter combinations in search for the best scores.

In the case of the LR the default hyperparameters yield the best results, while for the other algorithms the grid search shows that setting hyperparameters like the maximal tree depth and the number of estimators (here: DTs) in the ensemble resulted in better scores. The maximal tree depth limits the number of if-then-else decisions in the DTs, reducing the risk of overfitting the models. The best results are obtained by setting the parameters as follows: for the simple DT the maximal depth is set to 20, for the RF it is set to 40 and for the GB it is set to 15. The number of estimators is set to 600 and 400 for the RF and GB classifier, respectively. Additionally, the learning rate in the GB classifier is changed from 0.1 to 0.5, i.e., the fitting of the base estimators is accelerated slightly, without risking overfitting, by setting a higher number of estimators.

In addition to the CV score, we look at other scores that are often used for validating (binary) classifiers (Géron, 2019): the precision is the ratio of correct predictions out of all *inside magnetosphere* algorithm predictions; the recall or sensitivity is the ratio of correct predictions out of all true *inside* labels. For example, a recall of 0.95 for

Table 1. Final validation scores of different ML algorithms.

Score	Logistic Regression	Decision Tree	Random Forest	Gradient Boosting
CV	0.9633 ± 0.0012	0.9877 ± 0.0007	0.9939 ± 0.0005	0.9937 ± 0.0005
Precision	0.9606	0.9889	0.9939	0.9937
Recall	0.9722	0.9882	0.9938	0.9938
AUC	0.9944	0.9881	0.9998	0.9998

a classifier means that 95% of the data points inside the magnetosphere are predicted correctly.

To ascertain which model can distinguish best between the two classes, we also utilize the AUC (area under the curve) score. This score is derived as the integral from the receiver operating characteristic (ROC) curve, which describes the true positive rate (which is identical to the recall) as a function of the false positive rate (ratio of false predictions out of all true *outside* labels). For a purely random classifier, the AUC score would be 0.5, while a value of 1 would indicate a classifier perfectly distinguishing between the two classes.

The final validation scores for the algorithms are obtained by application of the trained models on the validation set TH07TV (after the setting of the hyperparameters according to the results of the grid search discussed above); They are depicted in Table 1. As can be seen, all algorithms share scores with values over 0.96 in all categories, and therefore could distinguish between the two classes and predict many magnetospheric data points correctly. By looking at the different scores in detail, it's clear that the DT performs better than the LR in regard to CV score, precision and recall; only in the AUC score LR shows higher values. Overall, the ensemble methods (RF and GB) perform even better than the simpler models, yielding nearly identical scores. The CV scores show the lowest standard deviation of $5{\cdot}10^{-4}$. Higher accuracies of 0.994 indicate a slightly better independence from the chosen training data. RF and GB also have precisions and recalls over 0.994 and AUC scores of 0.999. Thus, the ensemble methods are slightly better suited for the classification: they can distinguish very well between the two classes while also correctly predicting the labels in over 99% of the cases, matching the model comparison results of Nguyen et al. (2022a).

Finally, we compared the feature/input parameter importance of the RF and GB classifiers for the prediction of data points. The feature importance is a calculation of the relative contribution of each feature to the final decision, showing easily the influence of parameters to the model results. While the GB classifier mainly utilizes the ion density for its prediction, the RF uses many of the input parameters in its decision. This leads to the RF classifier being not as much affected by spurious density changes as the GB classifier; the latter tends to label density peaks erroneously as magnetosheath data, even if other observations suggested a different classification.

Thus, we decide to utilize the RF classifier to label the THEMIS data. We train it on our complete training dataset TH07T. The previous obtained scores are again verified by validating the RF classifier on the validation set TH07V.

3.2 Additional Threshold-based Corrections

Visual inspection of ~ 100 randomly selected intervals from 2007 to 2009 that were labelled with the Random Forest Classifier revealed some identification mistakes associated with foreshock phenomena or BS crossings. In addition, some mistakes were also found related to cold plasma observations deep inside the magnetosphere. To correct these mistakes, we use the following threshold-based label correction:

1. Southward IMF ($B_z \leq 0$ nT) and large ion velocities in AGSE x-direction ($v_x \leq -250$ km/s) should only be observed outside the dayside magnetosphere. If either (or both) of these criteria is fulfilled and if, in addition, ion densities above $n_{\mathrm{ion}} > 0.5\ \mathrm{cm}^{-3}$ are observed, then the associated points are relabelled as outside the magnetosphere.
2. High magnetic field magnitudes ($B > 150$ nT) and small deviations between the flux index and a high energy flux index ($F_{\mathrm{idx}}^{1e2,1e4} - F_{\mathrm{idx}}^{6.5e3,1e4} \leq 0.5$) should only be observed inside the magnetosphere. If either (or both) of these criteria is fulfilled and if, in addition, ion densities below $n_{\mathrm{ion}} < 0.75\ \mathrm{cm}^{-3}$ are observed, then the associated points are relabelled as inside the magnetosphere.

Roughly one percent of the labels have been corrected. The classification probability of these corrected labels is manually set to 0.85, indicating the correction.

We retrain our model on the gathered dataset of THEMIS data between 2007 and 2009 with corrected labels, trying to improve the classifier with these new labels. From here on, since directly adjacent points often share the same label, we choose a new data sampling rate of 12 s. Hence, we could accelerate the classification process without losing the accuracy of our model. Then we utilize the retrained Random Forest Classifier to label the remaining data up to 2022, while also applying the threshold-based label correction for 1% of the data.

3.3 Identification of Magnetopause Crossings

We search for MPCs by automatically identifying the times where labels change from one region to the other. We only count a label change as a MPC if at least two points before and after the change belong to the same region. That means a spacecraft has to be at least 24 s in a different region for a crossing to count.

The identification process results in an average of 13,164 MPCs per year. In total, 184,292 MPCs have been observed by the THEMIS spacecraft over the 15 years studied. These MPCs are collected into the dataset TH-MPC (Grimmich et al., 2023)

We calculate the deviation from the theoretical model stand-off distance ΔR_0 given by the SH98 model for each identified crossing

$$\Delta R_0 = R_{0,\mathrm{sc}} - R_{0,\mathrm{Shue}}, \tag{4}$$

$$R_{0,\mathrm{Shue}} = \left[10.22 + 1.29 \tanh\left(0.184\left(\frac{B_{z,\mathrm{IMF}}}{\mathrm{nT}} + 8.14\right)\right)\right]\left(\frac{p_{\mathrm{dyn}}}{\mathrm{nPa}}\right)^{-\frac{1}{6.6}}, \tag{5}$$

where equation (5) corresponds to equation (10) in SH98 and $B_{z,\mathrm{IMF}}$ and p_{dyn} are taken as the mean values in an event-preceding 8-minute interval from the solar wind OMNI dataset, taking the time delay from the bow shock to the magnetopause and the terminator into account. With definition (4), a negative ΔR_0 corresponds to a compression and a positive ΔR_0 to an expansion of the magnetopause to the spacecraft location.

We decided to project each of the observed MP location to its equivalent stand-off distance R_0. Otherwise, the flank MPCs could lead to a statistical bias towards higher values for MP position and model deviation, since the flank MP is naturally farther away from Earth and moves with higher amplitudes compared to the subsolar MP. We acknowledge that our method introduces errors associated with real flaring parameter differences to the SH98 model, which we discuss later in more detail.

In some cases ($\sim$11%) the stand-off distance and the deviation could not be calculated due to a lack of OMNI data for an entire interval, we have excluded the corresponding MPC entries from our database.

For each found MPC, we infer a crossing probability from the prediction probability $p_{\mathrm{RF}}(t)$ given by the RF classifier. The calculation is a weighted average of the probability of the 2 points before and after the jump in the labels:

$$p_{\mathrm{MPC}}(t_0) = \frac{1}{3}\left[p_{\mathrm{RF}}(t_0 - 12\text{ s}) + 0.5 p_{\mathrm{RF}}(t_0) + 0.5 p_{\mathrm{RF}}(t_0 + 12\text{ s}) + p_{\mathrm{RF}}(t_0 + 24\text{ s})\right]. \qquad (6)$$

The points are weighted with increasing time distance from the jump with 0.5 or 1 (see (6)), as the RF classifier predicts the labels with higher precision further away from the jump. The two points directly adjacent to the label change have the biggest prediction uncertainty and should contribute less to the probability calculation.

MPCs with low crossing probability are more likely misidentified or ambiguous. Thus, it is reasonable from here on to only use the roughly 75% of the database with high ($>$ 0.75) crossing probabilities (121,770 MPCs of TH-MPC). Additionally, as can be seen in Fig. 5, the MPC distributions with and without low crossing probability deviate essentially in count of events.

We point out that some misidentified crossings may still be left in the database, particularly in the high longitude region near the terminator, where a clean identification of crossings can be difficult, due to KHI-induced plasma mixing. Other misidentified crossings, which are still included in the database, are multiple crossings associated to a single extended magnetopause adjacent Low Latitude Boundary Layer (LLBL). This layer contains a mixture of magnetospheric and magnetosheath plasmas (e.g., Hasegawa, 2012), making a clear separation of the regions difficult and false multiple crossing detections more likely.

4 Magnetopause Crossing Statistics

Fig. 5 displays the distribution of all identified MPCs in the database over the stand-off distance, the deviation from the SH98 model in that distance and the latitude and longitude angles of the crossing positions. Separate distributions are shown for higher (> 0.75) and lower (< 0.75) crossing probabilities. In the top panel (a), the stand-off distance distribution is shown. We see a clear asymmetry around the maximum which lies roughly between 10.5 and 11 R_{E}: At 11.5 R_{E} a sharp decrease is seen, while for the smaller R_0 we see a smooth slope. The ΔR_0 distribution (panel (b)) indicates a tendency of the SH98 model to predict the MP a little nearer to Earth, as the maximum is at about $0.25 R_{\mathrm{E}}$. This may result from the fact that Shue et al. (1997, 1998) only used the innermost MPCs for fitting their model, while we do not restrict the database. Most of the MPCs are found between -1 and 1 R_{E} ($\sim$80%) which is consistent with reported SH98 model accuracies of $\sim \pm 1 R_{\mathrm{E}}$ (Case & Wild, 2013; Staples et al., 2020). As can be seen in the bottom two panels (c) and (d), the THEMIS orbits lead to MPC observations (1) being widely distributed in longitude ($|\lambda| < 90$) over the dayside and (2) being restricted in latitude to the near-equatorial region ($|\theta| < 30$).

If we compare the R_0-distribution with the distribution of the five THEMIS spacecraft dwell times at specific locations (Fig. 6), we see that the probes spent much more time in regions with $R_0 < 11\ R_{\mathrm{E}}$. Thus, the asymmetry in the MPC distribution results from this orbit bias which naturally leads to more MPCs at smaller stand-off distances. To compensate this orbit bias, we normalize the distributions, dividing the MPC count by the corresponding cumulative dwell time of all THEMIS spacecraft in each bin.

The normalization results can be seen in Fig. 7 showing the probability distribution of MPCs per hour of spacecraft observation time and also a comparison between

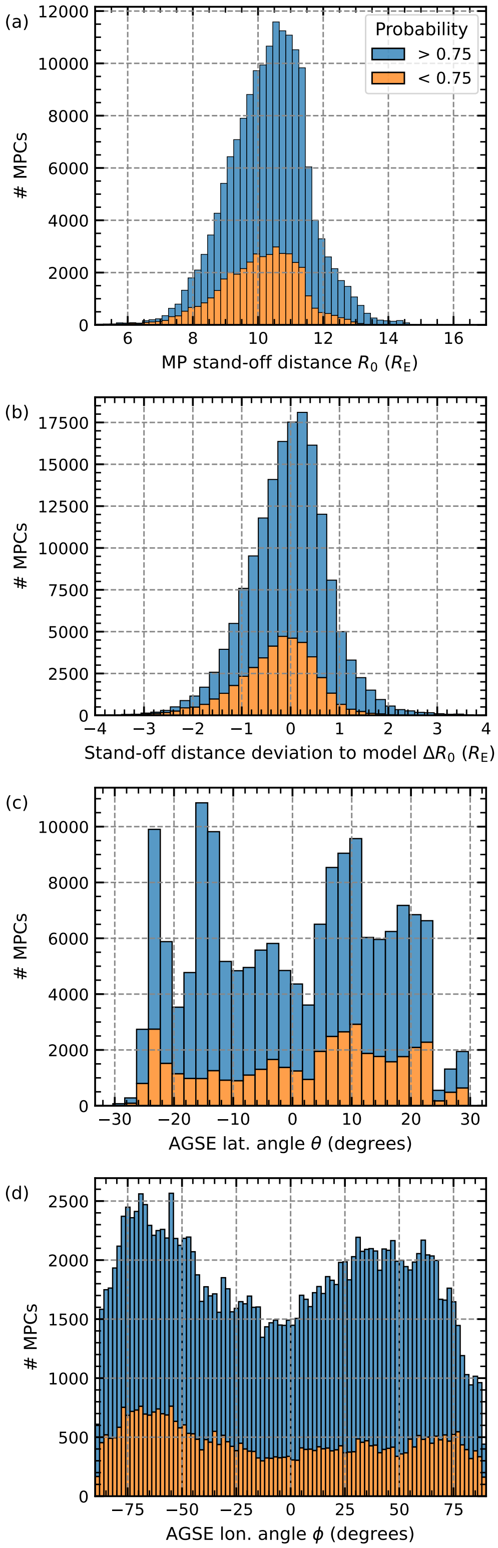

Figure 5. Distribution of detected MPCs, with detection probability ≤ 0.75 in blue and > 0.75 in orange. The panels show from top to bottom the stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, the latitude angle and the longitude angle of the respective MPCs in AGSE coordinates.

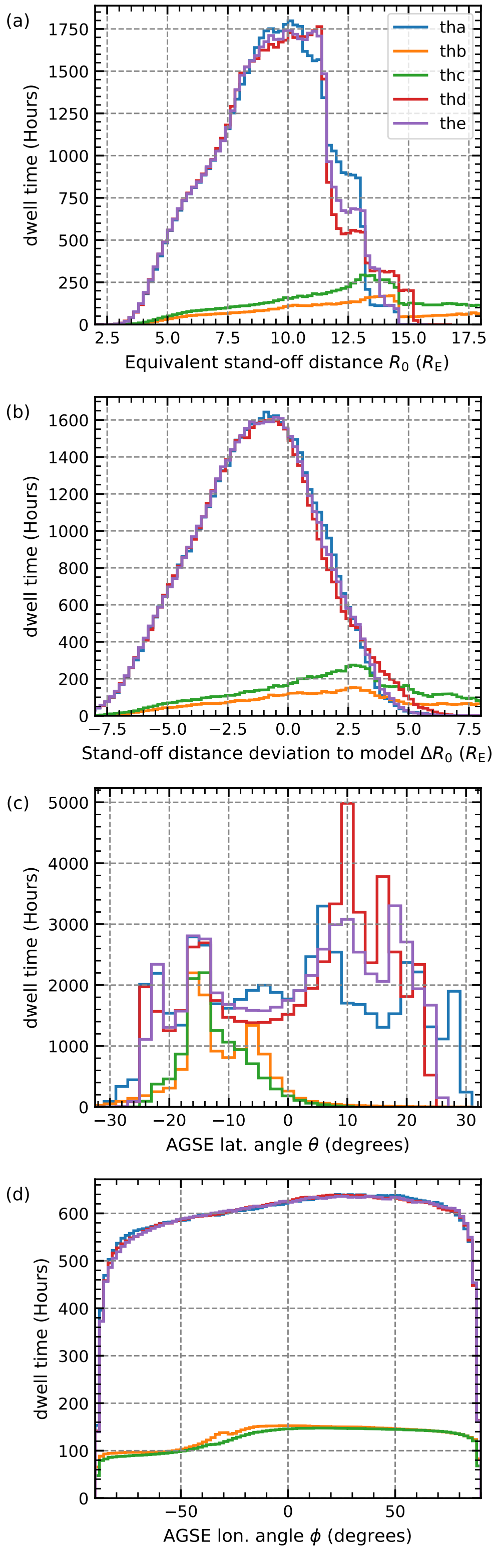

Figure 6. Dwell time distributions of the five THEMIS spacecraft with respect to the stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, the latitude angle and the longitude angle in AGSE coordinates (top to bottom).

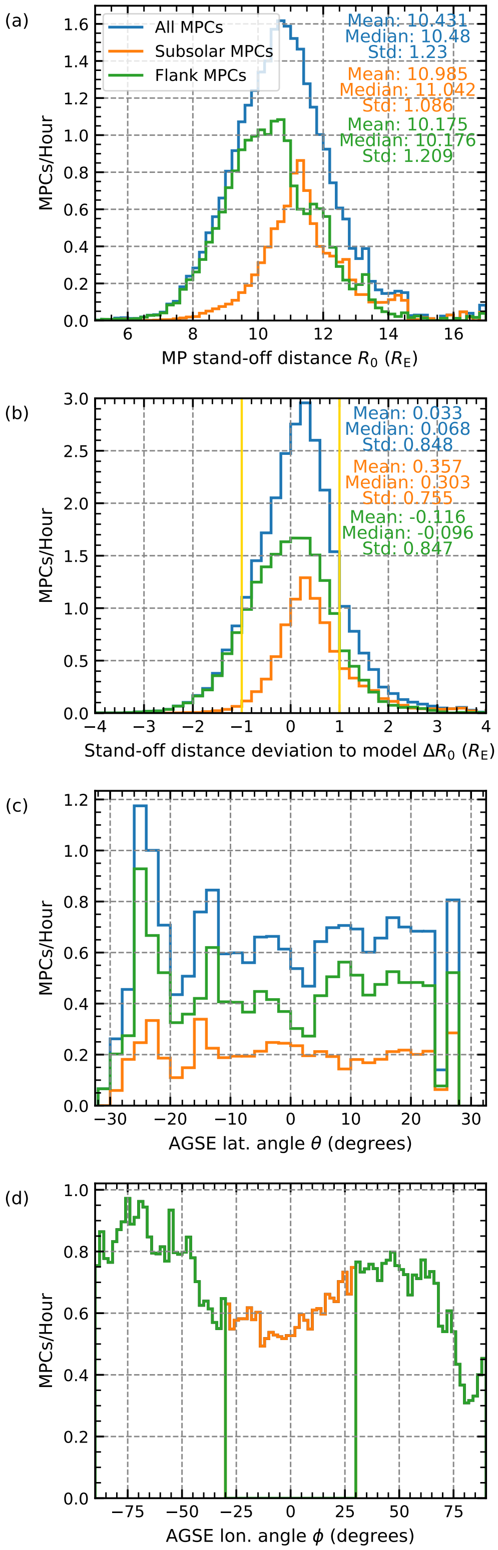

Figure 7. Histograms of the normalized distributions, showing crossing events per hour for each bin. The panels show the same variables as Fig. 3. The blue histogram depicts the hole dataset, while the orange and green histograms depicts the subsolar ($|\lambda| < 30$) and the flank ($|\lambda| > 30$) magnetopause subsets, respectively. The yellow line in panel 2 represents the uncertainty of the SH98 model.

subsolar ($|\lambda| \leq 30°$) and flank ($|\lambda| > 30°$) MPCs. The orbital bias in the stand-off distance (top panel) is no longer visible and the distribution is quasi symmetrical around 10.7 R_E indicated by the very similar mean and median values of the distribution. Interestingly, the subsolar MPCs occur slightly less frequently (0.86 MPCs/h) and the corresponding distribution is quite narrow in comparison to the broader flank MPCs distribution, which is centred around 10.5 R_E.

The SH98 model MP is dependent on the flaring parameter α and the stand-off distance R_0. On the day side, the flaring parameter has little influence on the MP position. Thus, adapting the SH98 MP to the MPC observations is achieved by changing the stand-off distance. At the flanks, motion of the MP results in variability of α. Since we fix the value of α with the prevalent solar wind conditions, all MP motion is attributed to changes in R_0, potentially leading to a broader distribution in this parameter (see Fig. 7a).

In the distribution of the deviations to the model (Fig. 7, panel (b)), the tendency to observe MPCs further away from Earth in comparison to model predictions is visible. Significant positive deviations from the SH98 model ($\Delta R_0 > 1$) result from expansions of the MP in the subsolar and flank regions while the significant negative deviations ($\Delta R_0 < -1$) result almost only from MP compressions in the flank regions.

Looking at the angular distributions of the MPCs, we find a notable asymmetry between the dawn and dusk sectors in the longitude distribution (bottom panel). The mean occurrence rate between -90° and -30° (dawn) is 0.79 MPCs/h while the rate between 30° and 90° (dusk) is 0.63 MPCs/h. In the subsolar sector the occurrence is in general lower than at the flank sectors (0.59 MPCs/h).

The MPCs are more or less equally distributed in latitude (panel (c)).

5 Solar Wind Statistics

5.1 Data Selection

The SH98 model magnetopause's location and shape are solely influenced by the solar wind dynamic pressure p_dyn and the IMF B_z-component. The model is nominally suitable to make predictions under extreme solar wind conditions which can lead to large deformations of the magnetopause (Shue et al., 1998). However, as shown in panel (b) of Fig. 7, we find numerous MPCs ($\sim$20%) outside the model uncertainties of ± 1 R_E occurring with rates ≤ 1.0 MPCs per hour.

About 7% of the MPCs in the database are classified as extreme deviations from the model stand-off distances, surpassing ± 1.5 R_E. Positive deviations correspond to magnetospheric expansions and negative deviations to magnetospheric compressions, in the following called expanded MPCs and compressed MPCs, respectively. From Fig. 7 we can infer that extreme expansions occur with rates ≤ 0.57 MPCs per hour and extreme compressions with rates ≤ 0.38 MPCs per hour.

These considerably deviating MPCs may be influenced by solar wind parameters that are not considered in the SH98 model. For this study, we associate each MPC from the high probability TH-MPC database with one set of solar wind parameters, comprised of the medians of the IMF magnitude B_IMF, the cone angle ϑ_cone between the Earth-Sun-line and the IMF vector, the clock angle ϑ_clock between the IMF B_y- and B_z-components, the solar wind velocity u_sw, the ion density n_ion, the ion temperature T_ion, the dynamic pressure p_dyn, the plasma β and the Alfvénic Mach number M_A, based on OMNI measurements form 8-minute intervals preceding each MPC.

5.2 Parameter Influence

To quantify the contribution of different solar wind parameters to the magnetopause distortions, we compare the whole distribution of the solar wind parameters from our OMNI dataset with the solar wind parameters associated with the TH-MPC database and the two extreme MPC subsets of expanded MPCs and compressed MPCs. We normalize each distribution individually by the total number of contributing data points.

The distributions with respect to B_{IMF}, $\vartheta_{\mathrm{cone}}$, $\vartheta_{\mathrm{clock}}$, u_{sw}, n_{ion}, T_{ion}, p_{dyn}, plasma β and M_A are shown in Figure 8. The OMNI data are shown in black and serves as reference. The solar wind data during the MPCs are shown in blue, while the orange and green lines display the distributions associated with extreme MPCs. The maxima and medians of the datasets are displayed as well, equally colour coded.

The solar wind data distributions (in black) agree nicely with results from previous studies (e.g., Plaschke et al., 2013; L. Q. Zhang et al., 2019; Larrodera & Cid, 2020; Ma et al., 2020). Furthermore, for all parameters we find an expected similarity in shape and maximum values between the blue and black distributions, as MPCs should be observed under all possible solar wind conditions over the long time range considered in this study. However, some of the distributions associated with extreme MPCs notably differ from the reference distributions, particularly with respect to $\vartheta_{\mathrm{cone}}$, u_{sw}, T_{ion} and M_A, indicating an influence of these parameters on the occurrence of extreme MP distortions. We compute the quotient of the distributions corresponding to the extreme MPCs with the reference solar wind distributions to indicate favourable occurrence conditions in the solar wind parameters. These favourable conditions are visible in quotient maxima above 1 and unfavourable conditions in minima under 1. In Fig. 9 these deviations from the reference distributions are displayed. The errors are computed using the mean detection rate of 15 MPCs per 1-hour interval as typical count error. In the following, we discuss the solar wind parameter distributions in the order of ascending influence on the extreme MPCs.

All clock angle distributions (Fig. 8F) show a double peak structure representing the known feature of the Parker spiral (e.g., L. Q. Zhang et al., 2019). In addition, we see small deviations in shape with respect to the reference solar wind distribution over all angles. Some clock angle orientations appear to be slightly more beneficial for the occurrence of extreme MPCs (see Fig. 9F). For example, the compressed MPCs show a tendency to occur under southward IMF conditions ($|\vartheta_{\mathrm{clock}}| \geq 100°$) and the distribution for the expanded MPCs deviates noticeable around $0°$, corresponding to occurrences during northward IMF. However, the positive deviations from 1, which can indicate favourable conditions seen in Fig. 9F are rather small in comparison to deviations in other parameters.

Although, the influence of the dynamic pressure on the magnetopause location should be captured by the SH98 model, we still see some subtle deviations in the distributions (panel J in Fig. 8) hinting at a further influence. We ignore the high peak at 0.3 nPa for the compressed MPCs in Fig. 9J as this large deviation results from only very few MPCs in this bin. Favourable conditions for extreme compressed MPCs are slightly higher pressures between 1.8 nPa and 3.5 nPa. The extreme expanded MPCs occur preferably under weaker pressures around 1.1 nPa. In both cases, however, the deviations in the distribution quotients found are less than 1.5. We conclude that the effect of dynamic pressure on extreme MPCs is already well captured by the SH98 model, as expected.

The distributions of plasma β (Fig. 8K) for extreme MPCs are slightly shifted with respect to the reference distribution. This shift is clearly visible in the maximum and the median values of the distributions. From Fig. 9K we infer that MP expansions occur more frequently for slightly higher β values between 2 and 5, and compressions are

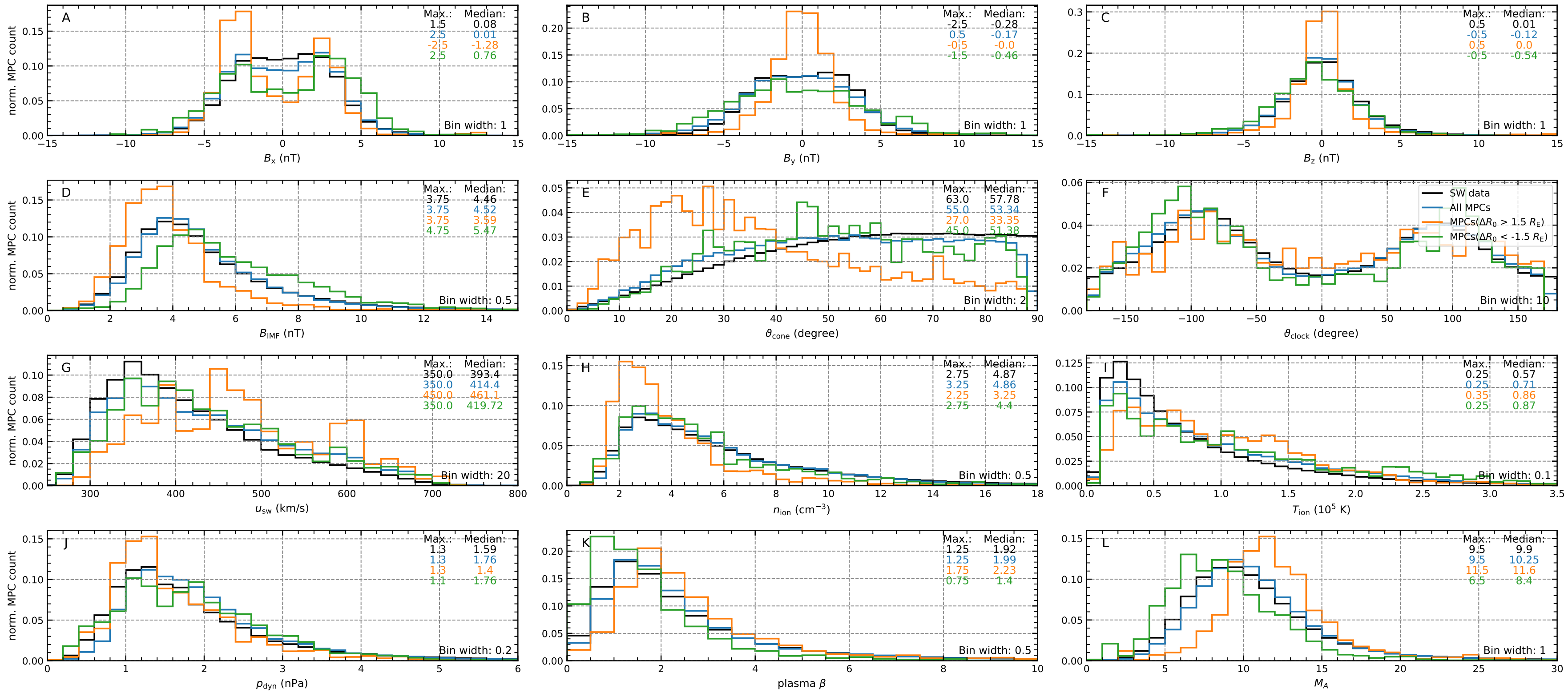

Figure 8. Probability distributions of different solar wind parameters. In each panel the whole solar wind distribution of a specific parameter (black lines) is compared to the parameter distribution pertaining to all MPCs (blue lines) and the extreme MPCs (orange and green lines, respectively). The maximum and the median values are indicated in each panel. All distributions are normalized by the total number of contributing data points.

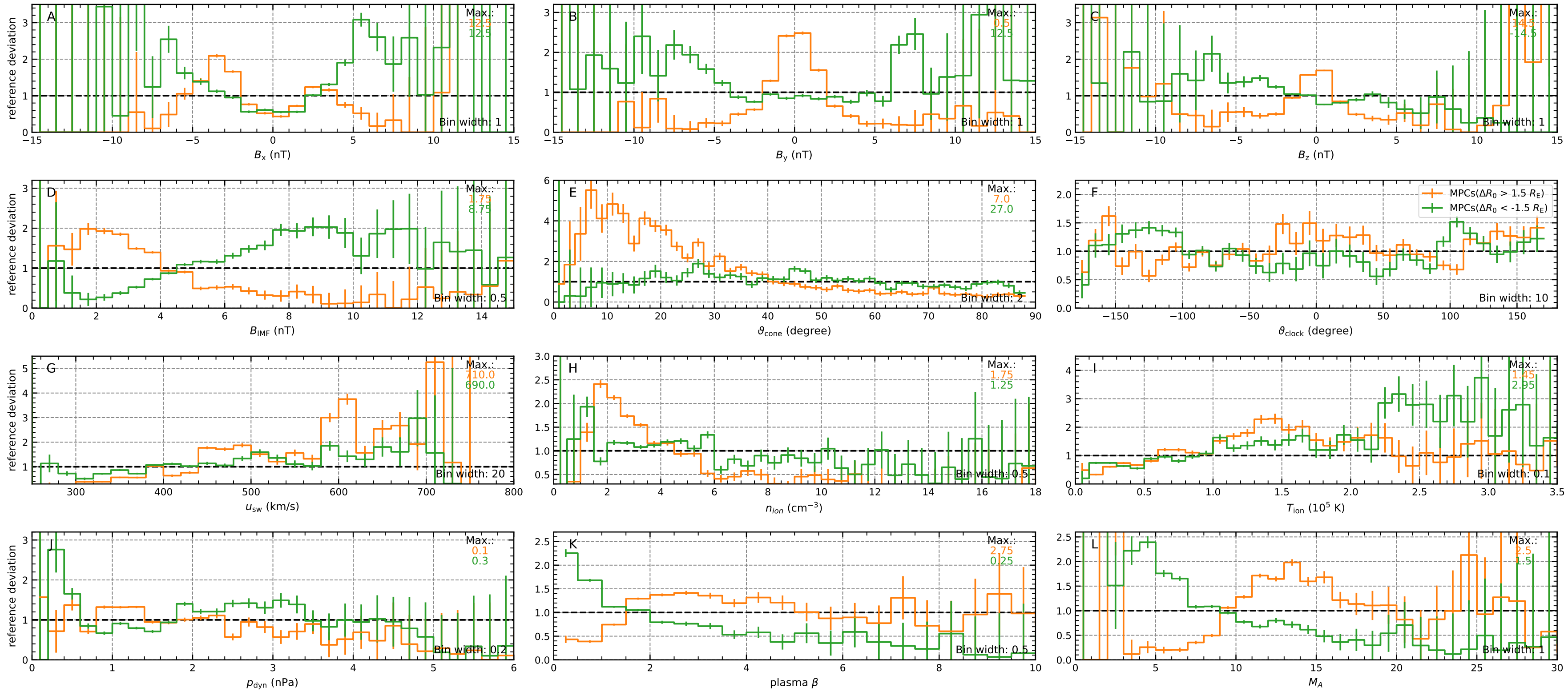

Figure 9. Deviations from the reference distributions (see Fig. 8) corresponding to extreme MPCs, calculated by dividing the distribution by the reference. The orange lines correspond to expanded MPCs an the green lines correspond to compressed MPCs.

more frequent for lower values below 1. Thus, higher/lower values lead to more frequent expansions/compressions.

In the ion density distributions (Fig. 8H), we find quite different deviations of the distributions for expanded and compressed MPCs. For the expanded MPCs, we can infer from Fig. 9H a clear tendency of higher occurrence rates between density values of 1.5 cm^{-3} and 3.5 cm^{-3}. For the compressed MPCs, we find one peak at $n_{\mathrm{ion}} = 1.25$ cm^{-3} which might be not reliable, as the bin contains only few MPCs. The other positive deviation for density values between 2 cm^{-3} and 6 cm^{-3} in the distribution quotient is very small.

Interestingly, all temperature distributions (Fig. 8I) share a common maximum around $3{\cdot}10^4$ K, but differ quite a lot in the median values. We find that the distributions for extreme MPCs are shifted to higher T_{ion}. Both compressed and expanded MPCs seem to occur more frequently in the temperature range between $1.0{\cdot}10^5$ K and $2.1{\cdot}10^5$ K (see Fig. 9I). Higher T_{ion} are favourable only for the compressed MPCs. However, in this temperature range we only observed very few MPCs.

From the three IMF components displayed in Fig. 8 and 9A - C we can also infer favourable conditions. Expanded MPCs occur more frequently for B_x around ±3 nT (negative values occur more often) and B_y around 0 nT. The compressed MPCs occur more often under even higher values of B_x around ±6 nT and for B_y values around ±7.5 nT. The influence of B_z is again well captured by the S98 model. However, similar to the influence of p_{dyn} we see some possible significant deviations regarding the compressed MPCs for strong negative values hinting at a favourable condition for these with B_z between -5 and -9 nT.

In panel D of Fig. 8 and 9 we can see that extreme expanded MPCs occur more frequently for smaller IMF magnitudes, with B_{IMF} between 1.5 nT and 4 nT. In contrast, the distribution of the extreme compressed MPCs is shifted to higher IMF magnitudes, indicating favourable conditions above 6 nT.

Fig. 8L, depicting the Alfvén Mach number distributions, shows obvious deviations between the reference and the extreme MPC distributions. The maxima and medians for the compressed and expanded MPCs deviate substantially from the reference, and we can clearly infer favourable conditions from Fig. 9L: For the expanded MPCs, we see the maximal occurrence rate at M_A =11.5 and favourable conditions of M_A between 11 and 16. For the compressed MPCs, we find the maximum at M_A =4.5 and favourable conditions of M_A between 3 and 7.

Both expanded and compressed MPCs seem to occur more frequently under high u_{sw} conditions (above 440 km/s). This trend is more clearly visible for the expanded MPCs (see Fig. 8G and 9G).

Lastly, we find a significant influence of $\vartheta_{\mathrm{cone}}$ on extreme expanded MPCs. Quasi-radial IMF conditions ($\vartheta_{\mathrm{cone}} < 35°$) clearly favour expanded MPCs (see panel E in Fig. 8 and 9). No similar feature can be seen with respect to the occurrence of compressed MPCs as a function of $\vartheta_{\mathrm{cone}}$. However, $\vartheta_{\mathrm{cone}}$ 25° and 30° could be a favourable condition for the compressed MPCs.

6 Discussion

In Fig. 7 (a), we find a quite symmetrical distribution of stand-off distances around 10.7 R_{E}, which can be regarded as typical (e.g. Baumjohann & Treumann, 1997). In comparison with stand-off distance predictions by the Shue et al. (1998) model (panel (b)), we find a slight tendency of the model to underestimate the stand-off distance, which probably results from the fact that Shue et al. (1998) only used the innermost crossings of MP encounters for fitting the model parameters.

In the longitude distribution of the MPCs (panel (d) of Fig. 7) we see a tendency to observe more MPCs at the magnetospheric flanks and a clear asymmetry between the occurrence rates in the dawn and dusk sectors. At the flanks, occurrences of KHI waves are likely (Taylor et al., 2012; Johnson et al., 2014) which should lead to frequent movement of the MP and more observations of MPCs. Additionally, as already mentioned, the Random Forest machine learning algorithm has some difficulties to clearly distinguish the magnetosphere and magnetosheath regions in case of thicker boundary layers leading to multiple crossing detections. We try to mitigate this problem by only studying MPCs with high crossing probabilities. Some remnants of this multiple MPCs might still be in the database, resulting in a tendency to observe more flank MPCs.

The dawn-dusk asymmetry is unlikely to be due to an orbital observation bias, as we have removed the bias corresponding to the spacecraft orbit using the spacecraft dwell time. Furthermore, our database is extensive enough that there should be no significant differences in the solar wind conditions prevalent during dawn and dusk MPCs. A dawn-dusk asymmetry in MPC occurrences has also been previously reported for the MPCs in the tail of the magnetosphere (e.g., Howe & Siscoe, 1972), and we can find the asymmetry in previous studies of the dayside MP. For example, Staples et al. (2020) used a threshold-based detection algorithm to study MPCs observed by THEMIS. Their MPC distributions (see Fig. A1) and ours are very similar, giving us confidence in our detection method using the Random Forest Classifier.

For Parker Spiral orientated IMF, KHIs predominately occur on the dawn flank (Henry et al., 2017), which might explain the previous statistical results. However other studies (e.g. Taylor et al., 2012) report on more frequent KHI occurrences at the dusk flank. Thus, KHIs might not be solely responsible for our observed slightly higher occurrence rate for MPCs at the dawn flank of the magnetosphere. Walsh et al. (2014) pointed out that the dawn flank has a thicker magnetopause boundary layer, which may give more weight to a possible explanation in terms of multiple MPC detections by the Random Forest Classifier. Another explanation for the asymmetry could be that the magnetopause moves more frequently in the dawn sector due to the thinner and more turbulent magnetosheath (Walsh et al., 2014). The foreshock will more often be located in this sector and excite more frequent MP movement, resulting in more frequent MPCs and therefore in higher occurrence rates. This foreshock effect on MP motion is also discussed in Russell et al. (1997) and is most likely related to the pressure variations associated with the foreshock region.

By comparing our database to the mentioned Staple et al. (2020) database, we can find out which explanation might be more reasonable. They looked at THEMIS data from 2007 to 2016 and only kept the innermost crossings of multiple MPCs in a 10-min interval. With the removal of multiple crossings, the above-mentioned higher occurrence rates due to KHIs or the detection method should not be visible in the distributions from Staples et al. (2020). However, as mentioned before, their database is subject to the dawn-dusk asymmetry in MPC occurrences. Hence, the reason for this asymmetry is more likely the more frequent occurrence of MP movement in the dawn sector possible linked to the foreshock or the more turbulent magnetosheath downstream of the quasi-parallel shock. Nevertheless, further investigations are necessary to fully understand this dawn-dusk asymmetry in MPC occurrences.

The statistical analysis above of the whole MPCs database is rather limited. Other studies like, e.g., Nguyen et al. (2022b) looked in more detail on the overall response of the magnetosphere to different solar wind parameters. Since we want to focus on the origin of the unusual MP locations, we only use the overall statistics to validate our identification method by comparing our results with those of previous studies. In doing so, we can see that some of our statistics regarding the whole MPC database (blue distributions in Fig. 8) are very similar to those found by Nguyen et al. (2022b). However, our statistical analysis revealed a clear asymmetry between the occurrence of MPC on the dawn and dusk flanks, which has not been clearly shown in other studies. In addition, we point out the influence of solar wind parameters on the occurrence of extreme MPCs. This is something that has rarely been discussed.

Let us now have a look at the roughly 7% of the identified MPCs that deviate drastically from the model predictions, that may not be immediately explained by changes in the solar wind dynamic pressure or the B_z-component of the IMF. From the comparison of the solar wind parameters during these extreme MPCs with the standard solar wind parameter distributions, we are able to infer some significant solar wind parameter influences on magnetopause location:

The most obvious influence pertains to the IMF cone angle, which controls the expansion of the magnetosphere as reported before (e.g., Slavin et al., 1996; Merka et al., 2003; Suvorova et al., 2010; Park et al., 2016; M. Wang et al., 2020). Under radial or quasi-radial IMF conditions, the dayside bow shock location is closer to Earth than on average, the magnetosheath thickness decreases, and the dayside magnetopause moves sunward. This happens in parts due to the establishment of a quasi-parallel foreshock in the subsolar region, which redistributes the dynamic pressure of the solar wind plasma and yields a lower magnetic pressure, affecting the magnetosphere. Additionally, the total plasma pressure is strongly modified in the bow shock crossing and distributed due to the flow diversion in the magnetosheath across the dayside magnetopause surface (Suvorova et al., 2010; Samsonov et al., 2012) leading to an expanding magnetopause to re-establish the pressure balance.

Extreme compressions might also occur under quasi-radial IMF conditions ($\vartheta_{\text{cone}} \approx 30°$). As Archer and Horbury (2013), Plaschke et al. (2013) and LaMoury et al. (2021) point out, HSJs occur more often under these conditions. Shue et al. (2009) and Archer et al. (2019) observed significant indentations of the magnetopause caused by a HSJ under radial IMF. Thus, the higher occurrence rate for compressions may be linked with such HSJs. However, the scale sizes of HSJ are small (Plaschke et al., 2020) leading only to a local indention of the MP. Thus, its much more likely that the MP response to the radial IMF on global scale is an expansion of the MP. This is also suggested by the distributions in Fig.9E, as the deviations for the expanded MPCs are clearly more significant in our database.

Substantial influences on extreme MP distortions stemming from the magnitude of the IMF, the plasma β and the Alfvén Mach number might in fact result from the same source: Extreme expansions of the magnetopause occur more frequently for small IMF magnitudes, i.e., values like the Alfvén velocity or the magnetic pressure are small as well. Naturally, small magnetic pressures and Alfvén velocities lead to higher plasma β and Alfvén Mach numbers, respectively, which are favourable conditions for extreme magnetospheric expansions as well. The expansions are possibly related to the higher particle reflection rates, leading to stronger kinetic energy dissipation at the bow shock under these condition (Winterhalter & Kivelson, 1988; Treumann, 2009), strengthening the foreshock region and thereby reducing the pressure on the dayside magnetosphere. On the contrary, the compressed MPCs occur more frequently for low Mach numbers and low plasma β, which result from higher IMF magnitudes. This might be connected to the fact, that the solar wind Mach number controls the magnetosheath plasma β. For low magnetosheath plasma β resulting from low solar wind Mach numbers, reconnection is more likely, leading to flux erosion and compression of the magnetosphere. Furthermore, we can infer that for magnetospheric compressions the magnetic pressure is dominant in the solar wind ($\beta < 1$), and for the magnetospheric expansions thermal pressure is dominant ($\beta > 1$).

Both in the velocity and in the temperature distributions (Fig. 8G and I), we identified a shift to higher values for extreme MPCs, especially noticeable for magnetospheric expansions. These higher values in the velocity probably lead to an increase in the velocity shear across the magnetopause triggering stronger KHIs and subsequentially exciting extreme oscillatory MP motion (Kavosi & Raeder, 2015). Additionally, Chu et al. (2017) and Vu et al. (2022) found that many different foreshock structures like FBs and HFAs were observed during radial IMF with solar wind velocities around 600 km/s. Generally, the favourable conditions of fast solar wind with large Alfvén Mach numbers and low cone angles for the extreme expansions coincide with favourable conditions for the occurrence of these foreshock transients (H. Zhang et al., 2022). These phenomena are characterized by hot tenuous plasma regions in the foreshock region, in which flow deflection and pressure reduction occur (Turner et al., 2013). On impact on the MP this pressure "hole" lead to an expansion followed by a compression of the magnetosphere (e.g., Sibeck et al., 1999; Turner et al., 2011; Archer et al., 2014, 2015). Thus, these foreshock transients might play an additional role in the extreme expansions and compressions of the magnetopause.

The B_x and B_y components of the IMF also seem to have major influences on the occurrence of extreme MPCs. Higher occurrence rates for expanded MPCs with B_y=0 and B_x strongly anti-sunwards indicate again a quasi-radial IMF as favourable conditions for such events. Therefore, the major deviations in the cone angle distribution might solely stem from the B_x influence. Results from simulations (Lu et al., 2013) suggest that for increasing IMF B_y the MP moves earthwards. This might be an important effect, as compressed MPCs occur more often for higher B_y values.

Weak but noticeable deviations in the clock angle distributions may stem from residual effects of the IMF B_z component, which is captured in the SH98 model. The possible favourable condition around 0° for expanded MPCs coincide with fact that the magnetopause stand-off distance is larger for northward IMF, when no reconnection and flux erosion is happening at the dayside. For southward IMF reconnection and flux erosion occur, driving the MP inwards and favouring observations of compressed MPCs under this condition.

The influence of the dynamic pressure is already prominently captured in the SH98 model. Therefore, the pressure effect would not appear as substantial deviation in our plots. As expected, similar behaviours can be found for the B_z component.

The deviation in the density distributions (Fig. 9H) for the compressed MPCs are also negligible, showing no significant favourable condition. In the distribution associated with the expanded MPCs we can see a minor preference for more tenuous solar wind plasma. Tenuous plasma causes a decrease in the thermal pressure, therefore reducing the total pressure impacting the magnetosphere.

Overall, we find more significant deviations from the reference distributions associated with the expanded MPCs, i.e., magnetospheric expansions are less well covered by the SH98 model. Therefore, we identify more favourable conditions for a extreme magnetospheric expansions. All influences on the magnetosphere presented here are additional effects besides the effects captured in the SH98 model. Not all of the effects might be solely responsible for the visible deviations in the distributions. To figure out which effects are most likely, further investigations are needed.

It should be emphasised that we use the SH98 model to identify extreme events deviating from the model. This model uses only two parameters for its prediction of the MP location and may miss previously reported influences on the MP steming from the dipole tilt or other IMF components (Lin et al., 2010; Liu et al., 2012; Lu et al., 2013). However, the SH98 model is widely used in the community and is able to predict the equatorial MP well on average. Therefore, we choose to use this model over others because of its simplicity. Of course, this could lead to a misinterpretation of our results, as the observed deviations could be due to the influence of the dipole tilt. To ensure that dipole tilt did not influence our results, we looked for seasonal effects in our distributions. Although we were able to find some variations between the winter and summer periods, there is no significant systematic influence of the dipole tilt on our results. This was expected, as Lu et al. (2013) indicated that dipole tilt does not affect the variation of equatorial MP.

7 Conclusion

In this study, the last 15 years of THEMIS observations have been used to build a very large MPC database. This database allows us to examine extreme MP distortions in detail in special case or statistical studies.

Our statistical study shows that parameters such as the Alfvén Mach number, the IMF cone angle and the ion velocity are responsible for quite frequent occurrences of extreme magnetopause distortions. Quasi-radial IMF conditions with a plasma $\beta > 1$, higher Alfvén Mach numbers and ion velocities above 450 km/s are favourable for magnetospheric expansions beyond the SH98 model predictions, while magnetospheric compressions are associated with more southward IMF conditions with plasma $\beta < 1$, lower Alfvén Mach numbers and IMF strengths above 5 nT.

The expansions of the magnetopause under high Mach number and velocity conditions are possible linked to KHIs and also foreshock transients, while other phenomena like magnetosheath jets might be responsible for some compressions. This could be studied further by comparing the observation times of such phenomena with our database. In sorting the extreme MPCs by possible origin mechanisms, we also hope to learn more about main drivers behind the extreme events.

Here we only study MPCs in low latitudes, observed on the dayside. With the utilization of CLUSTER data, we plan to expand our database to high latitudes (e.g., Panov et al., 2008), allowing for a comparison between the equatorial and more polar regions.

As some favourable conditions might stem from the same origin, it is also necessary to see if the favourable conditions are distinct from each other or more tightly connected. We plan to do this in a follow-up study.

With the upcoming SMILE mission, the shape and location of the MP will be directly inferred and linked to in-situ measurements of solar wind conditions. This will allow an immediate comparison with the results of this study and open the door for further investigations of extreme MP distortions.

Appendix A Threshold based MPC database

Staples et al. (2020) used the following criteria for the identification of MPCs on the dayside magnetosphere:

1. During a THEMIS crossing from the magnetosphere to the magnetosheath $\Delta B_{z,\mathrm{gsm}} < -0.6\frac{\mathrm{nT}}{\mathrm{s}}$ and $\Delta n_{\mathrm{ion}} > 0.08\frac{1}{\mathrm{cm}^3\mathrm{s}}$ should hold over the crossing. These criteria are reversed for crossings from the magnetosheath to the magnetosphere.
2. In average, $B_{z,\mathrm{gsm}} > 5$ nT and $n_{\mathrm{ion}} < 7$ cm^{-3} should hold for a 48-s interval within the magnetosphere before/after a possible event.
3. These two criteria must be satisfied in a 60-s-interval.

They completed their database with additional crossings identified by Plaschke, Glassmeier, Auster, Angelopoulos, et al. (2009), then removed duplicate crossings and reduced

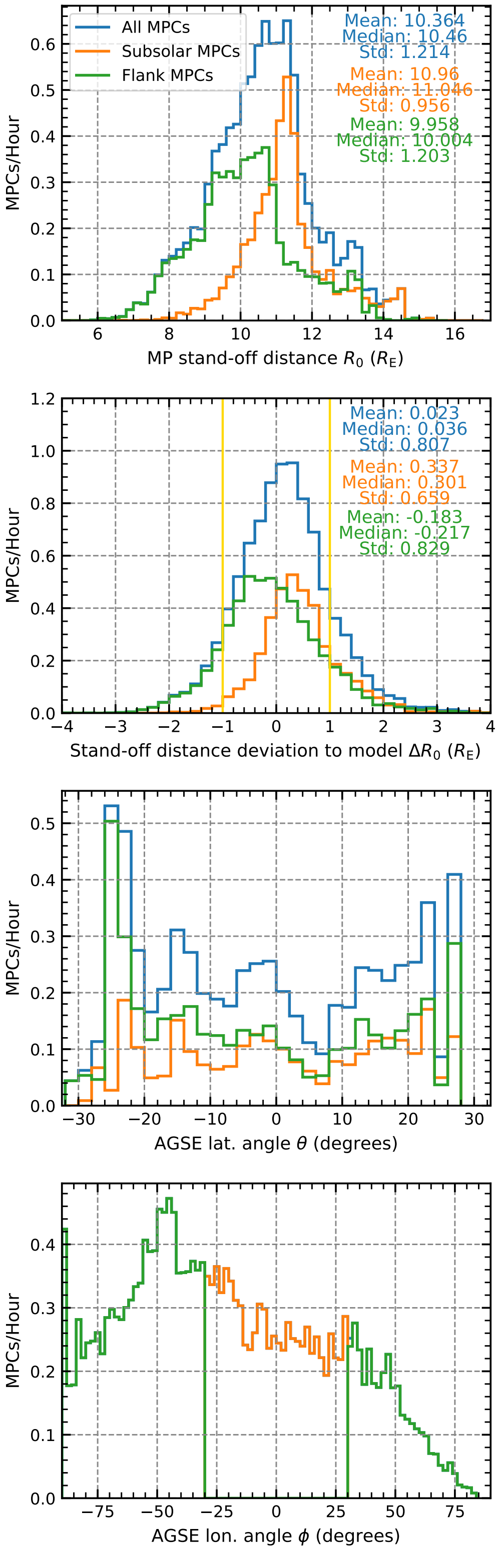

Figure A1. Histogram of normalized MPC distribution based on the database by Staples et al. (2020), showing crossing events per hour for each bin. The panels show from top to bottom: The stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, and the latitude and the longitude angles in AGSE coordinates. The blue histogram depict the hole dataset, while the orange and green histograms depicts the subsolar ($|\lambda| < 30$) and the flank ($|\lambda| > 30$) magnetopause subsets, respectively.

the database to the innermost crossings. Fig. A1 displays the distributions of their database identically normed as our database distributions.

Open Research

The magnetopause crossing event database constructed and used in this study is publicly available under `https://osf.io/b6kux/`, hosted by the Open Science Framework (OSF). To collect and plot data, we used the open source Python Space Physics Environment Data Analysis Software (pySPEDAS) which can be found here: `https://github.com/spedas/pyspedas`. THEMIS data can be retrieved from `http://themis.ssl.berkeley.edu/data/themis/` and OMNI data from the GSFC/SPDF OMNIWeb interface at `https://omniweb.gsfc.nasa.gov`. The machine learning task were performed with the scikit-learn Python library, from which we utilized the implementations of the different algorithms. The documentation can be found here: `https://scikit-learn.org/stable/supervised_learning.html#supervised-learning`.

Acknowledgments

We acknowledge NASA contract NAS5-02099 and V. Angelopoulos for use of data from the THEMIS Mission. Specifically: C. W. Carlson and J. P. McFadden for use of ESA data and K. H. Glassmeier, U. Auster and W. Baumjohann for the use of FGM data provided under the lead of the Technical University of Braunschweig and with financial support through the German Ministry for Economy and Technology and the German Center for Aviation and Space (DLR) under contract 50OC 0302. We thank J. King and N. Papitashvili of the National Space Science Data Center (NSSDC) in the NASA/GSFC for the use of the OMNI 2 database. FP and JZDM are supported by the Deutsches Zentrum für Luft- und Raumfahrt (DLR) under contract 50OC2201. DH was supported by the German Ministerium für Wirtschaft und Klimaschutz and the German Zentrum für Luft- und Raumfahrt (DLR) under contracts 50QW1501 and 50QW2202. MOA was supported by UKRI Stephen Hawking Fellowship EP/T01735X/1. The authors want to thank Nick Hatzigeorgiu and Eric Grimes for the ongoing development of PySPEDAS.

References

Angelopoulos, V. (2008, apr). The THEMIS Mission. *Space Sci. Rev.*, *141*(1-4), 5-34. doi: 10.1007/s11214-008-9336-1

Angelopoulos, V. (2011, dec). The ARTEMIS Mission. *Space Sci. Rev.*, *165*(1-4), 3-25. doi: 10.1007/s11214-010-9687-2

Archer, M. O., Hietala, H., Hartinger, M. D., Plaschke, F., & Angelopoulos, V. (2019, feb). Direct observations of a surface eigenmode of the dayside magnetopause. *Nature Communications*, *10*, 615. doi: 10.1038/s41467-018-08134-5

Archer, M. O., & Horbury, T. S. (2013, feb). Magnetosheath dynamic pressure enhancements: occurrence and typical properties. *Annales Geophysicae*, *31*(2), 319-331. doi: 10.5194/angeo-31-319-2013

Archer, M. O., Turner, D. L., Eastwood, J. P., Horbury, T. S., & Schwartz, S. J. (2014, oct). The role of pressure gradients in driving sunward magnetosheath flows and magnetopause motion. *Journal of Geophysical Research (Space Physics)*, *119*(10), 8117-8125. doi: 10.1002/2014JA020342

Archer, M. O., Turner, D. L., Eastwood, J. P., Schwartz, S. J., & Horbury, T. S. (2015, feb). Global impacts of a Foreshock Bubble: Magnetosheath, magnetopause and ground-based observations. *Planetary and Space Science*, *106*, 56-66. doi: 10.1016/j.pss.2014.11.026

Aubry, M. P., Russell, C. T., & Kivelson, M. G. (1970, jan). Inward motion of the magnetopause before a substorm. *Journal of Geophysical Research*, *75*(34), 7018. doi: 10.1029/JA075i034p07018

Auster, H. U., Glassmeier, K. H., Magnes, W., Aydogar, O., Baumjohann, W., Constantinescu, D., . . . Wiedemann, M. (2008, dec). The THEMIS Fluxgate Magnetometer. *Space Sci. Rev.*, *141*(1-4), 235-264. doi: 10.1007/s11214-008-9365-9

Baumjohann, W., & Treumann, R. (1997). *Basic Space Plasma Physics.* Imperial College Press.

Branduardi-Raymont, G., Wang, C., Escoubet, C., Adamovic, M., Agnolon, D., Berthomier, M., . . . Zhu, Z. (2018). SMILE definition study report. *European Space Agency, ESA/SCI*, *1*.

Breiman, L. (2001, jan). Random Forests. *Machine Learning*, *45*, 5-32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. S. (1984). *Classification and Regression Trees (1st Edition).* Chapman & Hall/CRC. doi: 10.1201/9781315139470

Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020, sep). Automatic classification of plasma regions in near-Earth space with supervised machine learning: application to Magnetospheric Multi Scale 2016-2019 observation. *Frontiers in Astronomy and Space Sciences*, *7*, 55. doi: 10.3389/fspas.2020.00055

Burch, J. L., Moore, T. E., Torbert, R. B., & Giles, B. L. (2016, mar). Magnetospheric Multiscale Overview and Science Objectives. *Space Sci. Rev.*, *199*(1-4), 5-21. doi: 10.1007/s11214-015-0164-9

Case, N. A., & Wild, J. A. (2013, oct). The location of the Earth's magnetopause: A comparison of modeled position and in situ Cluster data. *Journal of Geophysical Research (Space Physics)*, *118*(10), 6127-6135. doi: 10.1002/jgra.50572

Chao, J. K., Wu, D. J., Lin, C. H., Yang, Y. H., Wang, X. Y., Kessel, M., . . . Lepping, R. P. (2002, jan). Models for the Size and Shape of the Earth's Magnetopause and Bow Shock. In L.-H. Lyu (Ed.), *Space weather study using multipoint techniques* (p. 127).

Chu, C., Zhang, H., Sibeck, D., Otto, A., Zong, Q., Omidi, N., . . . Angelopoulos, V. (2017, mar). THEMIS satellite observations of hot flow anomalies at Earth's bow shock. *Annales Geophysicae*, *35*(3), 443-451. doi: 10.5194/angeo-35-443-2017

Cox, D. R., & Snell, E. J. (1970). *Analysis of Binary Data (2nd Edition).* Chapman & Hall/CRC. doi: 10.1201/9781315137391

Dušík, Š., Granko, G., Šafránková, J., Němeček, Z., & Jelínek, K. (2010, oct). IMF cone angle control of the magnetopause location: Statistical study. *Geophysical Research Letters*, *37*(19), L19103. doi: 10.1029/2010GL044965

Eastwood, J. P., Lucek, E. A., Mazelle, C., Meziane, K., Narita, Y., Pickett, J., & Treumann, R. A. (2005, jun). The Foreshock. *Sapce Sci. Rev.*, *118*(1-4), 41-94. doi: 10.1007/s11214-005-3824-3

Elphic, R. C. (1995, jan). Observations of Flux Transfer Events: A Review. *Geophysical Monograph Series*, *90*, 225. doi: 10.1029/GM090p0225

Escoubet, C. P., Fehringer, M., & Goldstein, M. (2001, oct). Introduction The Cluster mission. *Annales Geophysicae*, *19*, 1197-1200. doi: 10.5194/angeo-19-1197-2001

Fairfield, D. H. (1971, jan). Average and unusual locations of the Earth's magnetopause and bow shock. *Journal of Geophysical Research*, *76*(28), 6700. doi: 10.1029/JA076i028p06700

Fairfield, D. H., Baumjohann, W., Paschmann, G., Luehr, H., & Sibeck, D. G. (1990, apr). Upstream pressure variations associated with the bow shock and their effects on the magnetosphere. *Journal of Geophysical Research*, *95*(A4), 3773-3786. doi: 10.1029/JA095iA04p03773

Friedman, J. H. (2001, oct). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189 – 1232. Retrieved from `https://doi.org/10.1214/aos/1013203451` doi: 10.1214/aos/1013203451

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, Inc.

Grimmich, N., Plaschke, F., Archer, M. O., Heyner, D., Mieth, J. Z. D., Nakamura, R., & Sibeck, D. G. (2023). *Database: THEMIS magnetopause crossings between 2007 and mid-2022* [dataset]. Open Science Framework (OSF). Retrieved from `https://osf.io/b6kux/` doi: 10.17605/OSF.IO/B6KUX

Grygorov, K., Šafránková, J., Němeček, Z., Pi, G., Přech, L., & Urbář, J. (2017, nov). Shape of the equatorial magnetopause affected by the radial interplanetary magnetic field. *Planetary and Space Science*, *148*, 28-34. doi: 10.1016/j.pss.2017.09.011

Haaland, S., Hasegawa, H., Paschmann, G., Sonnerup, B., & Dunlop, M. (2021, aug). 20 Years of Cluster Observations: The Magnetopause. *Journal of Geophysical Research (Space Physics)*, *126*(8), e29362. doi: 10.1029/2021JA029362

Hasegawa, H. (2012, aug). Structure and Dynamics of the Magnetopause and Its Boundary Layers. *Monographs on Environment, Earth and Planets*, *1*(2), 71-119. doi: 10.5047/meep.2012.00102.0071

Henry, Z. W., Nykyri, K., Moore, T. W., Dimmock, A. P., & Ma, X. (2017, dec). On the Dawn-Dusk Asymmetry of the Kelvin-Helmholtz Instability Between 2007 and 2013. *Journal of Geophysical Research (Space Physics)*, *122*(12), 11,888-11,900. doi: 10.1002/2017JA024548

Howe, H. C., & Siscoe, G. L. (1972, jan). Magnetopause motions at lunar distance determined from the Explorer 35 Plasma Experiment. *Journal of Geophysical Research*, *77*(31), 6071. doi: 10.1029/JA077i031p06071

Jacobsen, K. S., Phan, T. D., Eastwood, J. P., Sibeck, D. G., Moen, J. I., Angelopoulos, V., . . . Fornaçon, K. H. (2009, aug). THEMIS observations of extreme magnetopause motion caused by a hot flow anomaly. *Journal of Geophysical Research (Space Physics)*, *114*(A8), A08210. doi: 10.1029/2008JA013873

Johnson, J. R., Wing, S., & Delamere, P. A. (2014, nov). Kelvin Helmholtz Instability in Planetary Magnetospheres. *Space Sci. Rev.*, *184*(1-4), 1-31. doi: 10.1007/s11214-014-0085-z

Kavosi, S., & Raeder, J. (2015, may). Ubiquity of Kelvin-Helmholtz waves at Earth's magnetopause. *Nature Communications*, *6*, 7019. doi: 10.1038/ncomms8019

King, J. H., & Papitashvili, N. E. (2005, feb). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *Journal of Geophysical Research (Space Physics)*, *110*(A2), A02104. doi: 10.1029/2004JA010649

Kuntz, K. D. (2019, jan). Solar wind charge exchange: an astrophysical nuisance. *The Astronomy and Astrophysics Review*, *27*(1), 1. doi: 10.1007/s00159-018-0114-0

LaMoury, A. T., Hietala, H., Plaschke, F., Vuorinen, L., & Eastwood, J. P. (2021, sep). Solar Wind Control of Magnetosheath Jet Formation and Propagation to the Magnetopause. *Journal of Geophysical Research (Space Physics)*, *126*(9), e29592. doi: 10.1029/2021JA029592

Larrodera, C., & Cid, C. (2020, mar). Bimodal distribution of the solar wind at 1 AU. *Astronomy and Astrophysics*, *635*, A44. doi: 10.1051/0004-6361/201937307

Lepping, R. P., Acũna, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., . . . Worley, E. M. (1995, feb). The Wind Magnetic Field Investigation. *Space Sci. Rev.*, *71*(1-4), 207-229. doi: 10.1007/BF00751330

Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010, apr). A three-dimensional asymmetric magnetopause model. *Journal of Geophysical Research (Space Physics)*, *115*(A4), A04207. doi: 10.1029/2009JA014235

Liu, Z. Q., Lu, J. Y., Kabin, K., Yang, Y. F., Zhao, M. X., & Cao, X. (2012, jul). Dipole tilt control of the magnetopause for southward IMF from global magnetohydrodynamic simulations. *Journal of Geophysical Research (Space Physics)*, *117*(A7), A07207. doi: 10.1029/2011JA017441

Liu, Z. Q., Lu, J. Y., Wang, C., Kabin, K., Zhao, J. S., Wang, M., ... Zhao, M. X. (2015, jul). A three-dimensional high Mach number asymmetric magnetopause model from global MHD simulation. *Journal of Geophysical Research (Space Physics)*, *120*(7), 5645-5666. doi: 10.1002/2014JA020961

Lu, J. Y., Liu, Z. Q., Kabin, K., Jing, H., Zhao, M. X., & Wang, Y. (2013, jun). The IMF dependence of the magnetopause from global MHD simulations. *Journal of Geophysical Research (Space Physics)*, *118*(6), 3113-3125. doi: 10.1002/jgra.50324

Ma, X., Nykyri, K., Dimmock, A., & Chu, C. (2020, oct). Statistical Study of Solar Wind, Magnetosheath, and Magnetotail Plasma and Field Properties: 12+ Years of THEMIS Observations and MHD Simulations. *Journal of Geophysical Research (Space Physics)*, *125*(10), e28209. doi: 10.1029/2020JA028209

McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., & Griffee, J. W. (1998, jul). Solar Wind Electron Proton Alpha Monitor (SWEPAM) for the Advanced Composition Explorer. *Space Sci. Rev.*, *86*, 563-612. doi: 10.1023/A:1005040232597

McFadden, J. P., Carlson, C. W., Larson, D., Ludlam, M., Abiad, R., Elliott, B., ... Angelopoulos, V. (2008, dec). The THEMIS ESA Plasma Instrument and In-flight Calibration. *Space Sci. Rev.*, *141*(1-4), 277-302. doi: 10.1007/s11214-008-9440-2

Merka, J., Szabo, A., Šafránková, J., & Němeček, Z. (2003, jul). Earth's bow shock and magnetopause in the case of a field-aligned upstream flow: Observation and model comparison. *Journal of Geophysical Research (Space Physics)*, *108*(A7), 1269. doi: 10.1029/2002JA009697

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2022a, jan). Massive Multi-Mission Statistical Study and Analytical Modeling of the Earth's Magnetopause: 1. A Gradient Boosting Based Automatic Detection of Near-Earth Regions. *Journal of Geophysical Research (Space Physics)*, *127*(1), e29773. doi: 10.1029/2021JA029773

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2022b, jan). Massive Multi-Mission Statistical Study and Analytical Modeling of the Earth's Magnetopause: 2. Shape and Location. *Journal of Geophysical Research (Space Physics)*, *127*(1), e29774. doi: 10.1029/2021JA029774

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2022c, jan). Massive Multi-Mission Statistical Study and Analytical Modeling of the Earth's Magnetopause: 3. An Asymmetric Non Indented Magnetopause Analytical Model. *Journal of Geophysical Research (Space Physics)*, *127*(1), e30112. doi: 10.1029/2021JA030112

Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., ... Gergin, E. (1995, feb). SWE, A Comprehensive Plasma Instrument for the Wind Spacecraft. *Space Sci. Rev.*, *71*(1-4), 55-77. doi: 10.1007/BF00751326

Olshevsky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzén, S., ... Markidis, S. (2021, oct). Automated Classification of Plasma Regions Using 3D Particle Energy Distributions. *Journal of Geophysical Research (Space Physics)*, *126*(10), e29620. doi: 10.1029/2021JA029620

Panov, E. V., Büchner, J., FräNz, M., Korth, A., Savin, S. P., RèMe, H., &

FornaçOn, K. H. (2008, jan). High-latitude Earth's magnetopause outside the cusp: Cluster observations. *Journal of Geophysical Research (Space Physics)*, *113*(A1), A01220. doi: 10.1029/2006JA012123

Park, J.-S., Shue, J.-H., Kim, K.-H., Pi, G., Němeček, Z., & Å afránková, J. (2016, jul). Global expansion of the dayside magnetopause for long-duration radial IMF events: Statistical study on GOES observations. *Journal of Geophysical Research (Space Physics)*, *121*(7), 6480-6492. doi: 10.1002/2016JA022772

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Plaschke, F., Glassmeier, K. H., Auster, H. U., Angelopoulos, V., Constantinescu, O. D., Fornaçon, K. H., ... Nakamura, R. (2009, jan). Statistical study of the magnetopause motion: First results from THEMIS. *Journal of Geophysical Research (Space Physics)*, *114*, A00C10. doi: 10.1029/2008JA013423

Plaschke, F., Glassmeier, K. H., Auster, H. U., Constantinescu, O. D., Magnes, W., Angelopoulos, V., ... McFadden, J. P. (2009, jan). Standing Alfvén waves at the magnetopause. *Geophysical Research Letters*, *36*(2), L02104. doi: 10.1029/2008GL036411

Plaschke, F., Hietala, H., & Angelopoulos, V. (2013, oct). Anti-sunward high-speed jets in the subsolar magnetosheath. *Annales Geophysicae*, *31*(10), 1877-1889. doi: 10.5194/angeo-31-1877-2013

Plaschke, F., Hietala, H., Archer, M., Blanco-Cano, X., Kajdič, P., Karlsson, T., ... Sibeck, D. (2018, aug). Jets Downstream of Collisionless Shocks. *Space Sci. Rev.*, *214*(5), 81. doi: 10.1007/s11214-018-0516-3

Plaschke, F., Hietala, H., & Vörös, Z. (2020, sep). Scale Sizes of Magnetosheath Jets. *Journal of Geophysical Research (Space Physics)*, *125*(9), e27962. doi: 10.1029/2020JA027962

Raab, W., Branduardi-Raymont, G., Wang, C., Dai, L., Donovan, E., Enno, G., ... Zheng, J. (2016, jul). SMILE: a joint ESA/CAS mission to investigate the interaction between the solar wind and Earth's magnetosphere. In J.-W. A. den Herder, T. Takahashi, & M. Bautz (Eds.), *Space telescopes and instrumentation 2016: Ultraviolet to gamma ray* (Vol. 9905, p. 990502). doi: 10.1117/12.2231984

Russell, C. T., Petrinec, S. M., Zhang, T. L., Song, P., & Kawano, H. (1997, jun). The effect of foreshock on the motion of the dayside magnetopause. *Geophysical Research Letters*, *24*(12), 1439-1441. doi: 10.1029/97GL01408

Safránková, J., Nemecek, Z., Dusík, S., Prech, L., Sibeck, D. G., & Borodkova, N. N. (2002, mar). The magnetopause shape and location: a comparison of the Interball and Geotail observations with models. *Annales Geophysicae*, *20*(3), 301-309. doi: 10.5194/angeo-20-301-2002

Samsonov, A. A., Němeček, Z., Šafránková, J., & Jelínek, K. (2012, may). Why does the subsolar magnetopause move sunward for radial interplanetary magnetic field? *Journal of Geophysical Research (Space Physics)*, *117*(A5), A05221. doi: 10.1029/2011JA017429

Shue, J. H., & Chao, J. K. (2013, jun). The role of enhanced thermal pressure in the earthward motion of the Earth's magnetopause. *Journal of Geophysical Research (Space Physics)*, *118*(6), 3017-3026. doi: 10.1002/jgra.50290

Shue, J. H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., & Singer, H. J. (1997, may). A new functional form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical Research*, *102*(A5), 9497-9512. doi: 10.1029/97JA00196

Shue, J. H., Chao, J. K., Song, P., McFadden, J. P., Suvorova, A., Angelopoulos, V., ... Plaschke, F. (2009, sep). Anomalous magnetosheath flows and distorted subsolar magnetopause for radial interplanetary magnetic fields. *Geophysical Research Letters*, *36*(18), L18112. doi: 10.1029/2009GL039842

Shue, J. H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., ... Kawano, H. (1998, aug). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research*, *103*(A8), 17691-17700. doi: 10.1029/98JA01103

Sibeck, D. G., Allen, R., Aryan, H., Bodewits, D., Brandt, P., Branduardi-Raymont, G., ... Wing, S. (2018, jun). Imaging Plasma Density Structures in the Soft X-Rays Generated by Solar Wind Charge Exchange with Neutrals. *Space Sci. Rev.*, *214*(4), 79. doi: 10.1007/s11214-018-0504-7

Sibeck, D. G., Borodkova, N. L., Schwartz, S. J., Owen, C. J., Kessel, R., Kokubun, S., ... Zastenker, G. N. (1999, mar). Comprehensive study of the magnetospheric response to a hot flow anomaly. *Journal of Geophysical Research*, *104*(A3), 4577-4594. doi: 10.1029/1998JA900021

Sibeck, D. G., Lopez, R. E., & Roelof, E. C. (1991, apr). Solar wind control of the magnetopause shape, location, and motion. *Journal of Geophysical Research*, *96*(A4), 5489-5495. doi: 10.1029/90JA02464

Slavin, J. A., Szabo, A., Peredo, M., Lepping, R. P., Fitzenreiter, R. J., Ogilvie, K. W., ... Steinberg, J. T. (1996, may). Near-simultaneous bow shock crossings by WIND and IMP 8 on December 1, 1994. *Geophysical Research Letters*, *23*(10), 1207-1210. doi: 10.1029/96GL01351

Smith, C. W., L'Heureux, J., Ness, N. F., Acuña, M. H., Burlaga, L. F., & Scheifele, J. (1998, jul). The ACE Magnetic Fields Experiment. *Space Sci. Rev*, *86*, 613-632. doi: 10.1023/A:1005092216668

Staples, F. A., Rae, I. J., Forsyth, C., Smith, A. R. A., Murphy, K. R., Raymer, K. M., ... Imber, S. M. (2020, apr). Do Statistical Models Capture the Dynamics of the Magnetopause During Sudden Magnetospheric Compressions? *Journal of Geophysical Research (Space Physics)*, *125*(4), e27289. doi: 10.1029/2019JA027289

Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., & Snow, F. (1998, jul). The Advanced Composition Explorer. *Space Sci. Rev.*, *86*, 1-22. doi: 10.1023/A:1005082526237

Suvorova, A. V., & Dmitriev, A. V. (2015, apr). Magnetopause inflation under radial IMF: Comparison of models. *Earth and Space Science*, *2*(4), 107-114. doi: 10.1002/2014EA000084

Suvorova, A. V., Shue, J. H., Dmitriev, A. V., Sibeck, D. G., McFadden, J. P., Hasegawa, H., ... Němeček, Z. (2010, oct). Magnetopause expansions for quasi-radial interplanetary magnetic field: THEMIS and Geotail observations. *Journal of Geophysical Research (Space Physics)*, *115*(A10), A10216. doi: 10.1029/2010JA015404

Tátrallyay, M., Erdös, G., Németh, Z., Verigin, M. I., & Vennerstrom, S. (2012, dec). Multispacecraft observations of the terrestrial bow shock and magnetopause during extreme solar wind disturbances. *Annales Geophysicae*, *30*(12), 1675-1692. doi: 10.5194/angeo-30-1675-2012

Taylor, M. G. G. T., Hasegawa, H., Lavraud, B., Phan, T., Escoubet, C. P., Dunlop, M. W., ... Wild, J. A. (2012, jun). Spatial distribution of rolled up Kelvin-Helmholtz vortices at Earth's dayside and flank magnetopause. *Annales Geophysicae*, *30*(6), 1025-1035. doi: 10.5194/angeo-30-1025-2012

Treumann, R. A. (2009, dec). Fundamentals of collisionless shocks for astrophysical application, 1. Non-relativistic shocks. *The Astronomy and Astrophysics Review*, *17*(4), 409-535. doi: 10.1007/s00159-009-0024-2

Turner, D. L., Eriksson, S., Phan, T. D., Angelopoulos, V., Tu, W., Liu, W., ... Glassmeier, K. H. (2011, apr). Multispacecraft observations of a foreshock-induced magnetopause disturbance exhibiting distinct plasma flows and an intense density compression. *Journal of Geophysical Research (Space Physics)*, *116*(A4), A04230. doi: 10.1029/2010JA015668

Turner, D. L., Omidi, N., Sibeck, D. G., & Angelopoulos, V. (2013, apr). First observations of foreshock bubbles upstream of Earth's bow shock: Characteristics and comparisons to HFAs. *Journal of Geophysical Research (Space Physics)*, *118*(4), 1552-1570. doi: 10.1002/jgra.50198

Vu, A., Liu, T. Z., Zhang, H., & Pollock, C. (2022, feb). Statistical Study of Foreshock Bubbles, Hot Flow Anomalies, and Spontaneous Hot Flow Anomalies and Their Substructures Observed by MMS. *Journal of Geophysical Research (Space Physics)*, *127*(2), e2021JA030029. doi: 10.1029/2021JA030029

Walsh, A. P., Haaland, S., Forsyth, C., Keesee, A. M., Kissinger, J., Li, K., … Taylor, M. G. G. T. (2014, jul). Dawn-dusk asymmetries in the coupled solar wind-magnetosphere-ionosphere system: a review. *Annales Geophysicae*, *32*(7), 705-737. doi: 10.5194/angeo-32-705-2014

Wang, C., & Sun, T. (2022, dec). Methods to derive the magnetopause from soft X-ray images by the SMILE mission. *Geoscience Letters*, *9*(1), 30. doi: 10.1186/s40562-022-00240-z

Wang, M., Lu, J. Y., Kabin, K., Yuan, H. Z., Zhou, Y., & Guan, H. Y. (2020, may). Influence of the Interplanetary Magnetic Field Cone Angle on the Geometry of Bow Shocks. *The Astronomical Journal*, *159*(5), 227. doi: 10.3847/1538-3881/ab86a7

Winterhalter, D., & Kivelson, M. G. (1988, sep). Observations of the Earth's bow shock under high Mach number/high plasma beta solar wind conditions. *Geophysical Research Letters*, *15*(10), 1161-1164. doi: 10.1029/GL015i010p01161

Zhang, H., Zong, Q., Connor, H., Delamere, P., Facskó, G., Han, D., … Yao, S. (2022, aug). Dayside Transient Phenomena and Their Impact on the Magnetosphere and Ionosphere. *Space Sci. Rev.*, *218*(5), 40. doi: 10.1007/s11214-021-00865-0

Zhang, L. Q., Wang, C., Wang, J. Y., & Lui, A. T. Y. (2019, may). Statistical properties of the IMF clock angle in the solar wind with northward and southward interplanetary magnetic field based on ACE observation from 1998 to 2009: Dependence on the temporal scale of the solar wind. *Advances in Space Research*, *63*(10), 3077-3087. doi: 10.1016/j.asr.2019.01.023